

EXHIBIT A

MEALEY'S™ LITIGATION REPORT
Discovery

Predictive Coding: A Primer

by
Amy Jane Longo, Esq.
and
Usama Kahf, Esq.

O'Melveny & Myers LLP
Los Angeles, California

**A commentary article
reprinted from the
March 2013 issue of
Mealey's Litigation
Report: Discovery**

Commentary

Predictive Coding: A Primer

By
Amy Jane Longo
and
Usama Kahf

[Editor's Note: Amy Longo is a partner at O'Melveny & Myers LLP and a member of the Firm's Financial Services and Electronic Discovery and Document Retention Practices. Usama Kahf is an associate and a member of the Firm's Labor and Employment Practice. Both are resident in O'Melveny's Los Angeles office. The opinions expressed in this article do not necessarily reflect the views of O'Melveny or its clients, and should not be relied upon as legal advice. Copyright © 2013 by Amy Jane Longo and Usama Kahf. Responses to this commentary are welcome.]

Having gained judicial approval — or acknowledgement — in fewer than a handful of cases, the method for collecting and reviewing electronic documents for discovery known as “predictive coding” appears to be “trending” — to borrow a term from social media culture. One way to understand the emergence of predictive coding — which is referred to alternately as computer-assisted review, technology-assisted review or intelligent review — is as an answer to the many critiques levied over the years by courts, litigants, and legal scholars about the adequacy of the “search terms” method of culling through repositories of electronically stored information (ESI) for relevant discoverable evidence.¹ With the exponential increase in ESI discovery over the last decade, courts regularly address challenges to the sufficiency of producing parties’ efforts to search for and collect relevant ESI.² Discovery disputes abound over the search terms utilized, number of custodians, types of repositories searched, and method of collection.³

To satisfy a party’s obligations in responding to discovery requests, the method the party uses to search, collect, and review ESI must be defensible as reasonable, if challenged.⁴ Importantly, while cost of discovery is

certainly an important consideration for all parties, the legal adequacy of any search and review method must turn on its precision and accuracy in culling relevant documents and appropriately determining their responsiveness to discovery requests, not whether the method of choice provides costs savings relative to other methods.

Predictive coding involves “training” a computer (by way of an electronic discovery software or review tool) to recognize and identify the documents in a review set that are relevant and/or responsive to discovery requests. The software “learns” how to code from the case team’s attorneys, as it tracks their review decisions (i.e., which tags are checked) and uses mathematical algorithms to predict based on the contents and characteristics of documents which tags the attorneys would have checked.

Predictive coding first received judicial endorsement as a reliable discovery review method in a well-publicized article on October 1, 2011, by Magistrate Judge Andrew Peck of the Southern District of New York.⁵ Noting that at that point no court had approved of predictive coding, Judge Peck wrote, “Until there is a judicial opinion approving (or even critiquing) the use of predictive coding, counsel will just have to rely on this article as a sign of judicial approval. In my opinion, computer-assisted coding should be used in those cases where it will help ‘secure the just, speedy, and inexpensive’ determination of cases in our ediscovery world.”⁶ Only a few months later, the defendants in a gender discrimination case before Judge Peck, *Da Silva Moore v. Publicis Groupe & MSL Group*,⁷ sought his approval to use predictive coding without objection from plaintiffs to this method other than to the

wording of the proposed stipulation and certain aspects of Judge Peck's ruling. His ruling, which was later approved by the district court,⁸ became the first judicial opinion in which a court approved the use of predictive coding in searching for and reviewing ESI.

As Judge Peck explained, both in his article and in *Da Silva Moore*, the traditional method of using keywords to search through repositories of ESI for relevant documents has many flaws. Specifically, requesting parties often simply guess "which keywords might produce evidence to support its case without having much, if any, knowledge of the responding party's 'cards' (i.e., the terminology used by the responding party's custodians) ... [and] the responding party's counsel often does not know what is in its own client's 'cards.'"⁹ Poorly designed search terms generally do not benefit any party (whether the requestor or the recipient), as the terms tend to be both over-inclusive (returning many documents that are not responsive) and under-inclusive (failing to identify responsive documents).

As a potential alternative, predictive coding enables the computer, with human input and coaching, to determine relevance based on sophisticated algorithms and its observations of the human reviewer's review decisions. One important aspect of predictive coding is that in order to properly "train" the computer on what makes documents relevant (i.e., in order for predictive coding to be successfully used), a more senior attorney familiar with both the law and facts of the case — rather than junior attorneys or contract reviewers — must review and code a "seed set" of documents. This may require an initial investment of time by the more senior attorney, but the effectiveness of predictive coding highly depends on the reliability of the "seed set." As the attorney codes the seed set, the computer identifies characteristics of the coded documents (including, for example, various metadata fields) and begins to associate those characteristics with how the attorney coded the documents. The attorney continues to review the "seed set" until the computer learns enough from the attorney about which documents are relevant. This happens when the computer starts to predict the attorney's coding with increasing rates of precision and accuracy. Soon, the computer's predictions begin to correlate with the attorney's coding. At this point, review of the "seed set" is complete, and the attorney can be confident that the computer can now take over the initial coding of documents and be just as accurate and reliable.¹⁰

In *Da Silva Moore*, Judge Peck concluded that use of predictive coding software as agreed to by the parties in their ESI discovery protocol is more adequate to satisfy discovery obligations than keyword searching.¹¹ Judge Peck acknowledged that predictive coding is not perfect and cannot be expected to be, but neither is using search terms.¹² He concluded with the following message to practitioners:

What the Bar should take away from this Opinion is that computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review. Counsel no longer have to worry about being the "first" or "guinea pig" for judicial acceptance of computer-assisted review. As with keywords or any other technological solution to ediscovery, counsel must design an appropriate process, including use of available technology, with appropriate quality control testing, to review and produce relevant ESI while adhering to Rule 1 and Rule 26(b)(2)(C) proportionality. Computer-assisted review now can be considered judicially-approved for use in appropriate cases.¹³

Since Judge Peck's endorsement of predictive coding in *Da Silva Moore*, several state and federal courts have discussed, and some have approved, use of this method in ESI discovery. For example, in *National Day Laborer Organizing Network v. U.S. Immigration and Customs Enforcement Agency*,¹⁴ a Freedom of Information Act case, Judge Shira Scheindlin of the Southern District of New York pointed to predictive coding as an example of what she called "emerging best practices" for proper search, collection, and review of ESI. Judge Scheindlin noted that "parties can (and frequently should) rely on latent semantic indexing, statistical probability models, and machine learning tools to find responsive documents. Through iterative learning, these methods (known as 'computer-assisted' or 'predictive' coding) allow humans to teach computers what documents are and are not responsive to a particular ... discovery request and they can significantly increase the effectiveness and efficiency of searches."¹⁵

MEALEY'S LITIGATION REPORT: Discovery

As of the end of 2012, besides this positive mention in *National Day Laborer Organizing Network*, predictive coding has been approved by one additional federal court and two state courts, and another federal court has held an evidentiary hearing on the method's adequacy.

First, in a multidistrict case involving allegations that the diabetes medication Actos increases users' risk of developing bladder cancer, *In re Actos (Pioglitazone) Products Liability Litigation*,¹⁶ U.S. Magistrate Judge Hanna Doherty of the Western District of Louisiana issued a case management order regarding discovery of ESI with comprehensive instructions to guide the parties' use of technology-assisted review. The order, which was stipulated to by the parties, provided specific directions on how the parties should consider and treat data sources, custodians, costs, and format of production, among other discovery questions. Notably, the order included a "Search Methodology Proof of Concept" governing the parties' use of technology-assisted review tools to search for, collect, and review ESI. The order stated that the parties "agree to meet and confer regarding the use of advanced analytics" as a "document identification mechanism for the review and production of ... data," as well as to select four key custodians whose email will be used to create an initial seed set, after which three experts would "train" the software on coding documents based on relevance. As a check on the reliability of this method, the order directed both parties to collaborate to train the software and to mutually decide upon the appropriate threshold of relevance, with the right to seek input from the court in case of a dispute. The order also provided that after sufficient training of the software, the documents coded by the software must be randomly sampled for quality control, and the defendants would retain the right to manually review documents prior to production for relevance, confidentiality, and privilege.

Second, in *Global Aerospace v. Landow Aviation*,¹⁷ the Virginia Circuit Court similarly approved a predictive coding protocol after defendants moved for a protective order due to the volume of data sought by plaintiffs. Relying on *Da Silva Moore*, defendants successfully argued that predictive coding was not only significantly

less expensive than manual review and keyword searches, but also significantly more reliable in identifying relevant documents.

Third, a judge in the Delaware Court of Chancery surprised both plaintiffs and defendants in *EORHB v. HOA Holdings LLC*¹⁸ when he issued an order from the bench that required both sides to use predictive coding with the same vendor for ESI discovery (or otherwise to "show cause why this is not a case where predictive coding is the way to go"). This appears to be the first time a judge has required both parties in a case to use predictive coding when neither requested it.

Finally, in *Kleen Products LLC v. Packaging Corp. of America*,¹⁹ the plaintiffs challenged the defendants' proposed Boolean search methodology as likely to find less than 25% of responsive documents and asserted that content-based advance analytics search (i.e., predictive coding) would identify more than 70% of responsive documents at no greater cost.²⁰ Plaintiffs argued that predictive coding would "not focus on matching words but instead on identifying relevant concepts out of the documents," and would "provide a richer, substantially more accurate return than Boolean searches."²¹ Plaintiffs criticized Defendants' Boolean keyword search as per se "subject to the inadequacies and flaws inherent when keywords are used to identify responsive documents."²² Defendants defended their search method on grounds that their quality control processes will ensure a "degree of accuracy" in line with industry standards.²³ They also contended that predictive coding would involve additional costs and burdens not contemplated by the Federal Rules, local rules or case law.²⁴ The *Kleen Products* court did not directly rule on the propriety of predictive coding because after an evidentiary hearing and five months of meeting and conferring, plaintiffs agreed to withdraw their demand that defendants use predictive coding.²⁵

Predictive coding technology continues to evolve rapidly, as concerns over its mechanics and reliability are addressed. One critique of this method is based on the notion that there is no substitute for an attorney's judgment, as computer software may potentially replace humans in document reviews. A balanced approach that incorporates attorney judgment and discretion along the way may alleviate some of these

concerns. Until such a method is tried and tested, predictive coding will remain on the cutting edge.

Endnotes

1. See, e.g., *Chura v. Delmar Gardens of Lenexa, Inc.*, 2012 U.S. Dist. LEXIS 36893, *33-35 (D. Kan. Mar. 20, 2012) (ordering an evidentiary hearing to explore the sufficiency of the defendant's search for responsive ESI where it failed to produce emails and other electronic documents and allegedly failed to do more than run a search for terms in an email program on one computer); *Custom Hardware Eng'g & Consulting v. Dowell*, 2012 U.S. Dist. LEXIS 146, *7-8 (E.D. Mo. Jan. 3, 2012) (discussing the problems that arise from keyword search methodologies, particularly that the word choices may be arbitrary and unable to reach the relevant information).
2. E.g., *Williams Mullen v. U.S. Army Criminal Investigation Command*, 2012 U.S. Dist. LEXIS 93977, *12 (E.D. Va. July 6, 2012) (outlining the standard for whether a party's search for responsive documents was adequate, including whether it was "reasonably calculated to uncover all relevant documents"); *William A. Gross Const. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134, 136 (S.D.N.Y. 2009) ("[W]here counsel are using keyword searches for retrieval of ESI, they at a minimum must carefully craft the appropriate keywords, with input from the ESI's custodians as to the words and abbreviations they use, and the proposed methodology must be quality control tested to assure accuracy in retrieval and elimination of 'false positives.'").
3. See, e.g., *S2 Automation LLC v. Micron Tech., Inc.*, 2012 U.S. Dist. LEXIS 120097, *1 (D.N.M. Aug. 14, 2012) (addressing whether a party must produce ESI in the format requested, whether it must separately produce metadata, and whether it must disclose the search strategy used for production of documents); *Orillaneda v. French Culinary Inst.*, 2011 U.S. Dist. LEXIS 105793, *27 (S.D.N.Y. Sept. 19, 2011) (denying a plaintiff's requests for discovery of how defendant searched and maintained its information systems where the plaintiff failed to present any specific reason to believe defendant's responses to electronic discovery requests were deficient).
4. See *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 262 (D. Md. 2008) ("Selection of the appropriate search and information retrieval technique requires careful advance planning by persons qualified to design effective search methodology. The implementation of the methodology selected should be tested for quality assurance; and the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented. In this regard, compliance with the Sedona Conference Best Practices for use of search and information retrieval will go a long way towards convincing the court that the method chosen was reasonable and reliable."). See also William W. Belt, Dennis R. Kiker & Daryl E. Shetterly, *Technology-Assisted Document Review: Is It Defensible?*, XVIII RICH. J. L. & TECH. 10 (2012), available at <http://jolt.richmond.edu/v18i3/article10.pdf>.
5. Andrew Peck, *Search, Forward*, LAW TECHNOLOGY NEWS (Oct. 1, 2011) ("*Search, Forward*") (citing Maura Grossman & Gordon Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, XVII RICH. J. L. & TECH. 11 (2011), available at <http://jolt.richmond.edu/v17i3/article11.pdf>).
6. *Id.*
7. 2012 U.S. Dist. LEXIS 23350 (S.D.N.Y. Feb. 24, 2012) (M.J. Peck).
8. *Da Silva Moore v. Publicis Groupe & MSL Group*, 2012 U.S. Dist. LEXIS 58742 (S.D.N.Y. Apr. 26, 2012) (approving Judge Peck's endorsement of predictive coding).
9. *Search, Forward*.
10. *Id.*
11. 2012 U.S. Dist. LEXIS 23350, *35 (S.D.N.Y. Feb. 24, 2012) (M.J. Peck).
12. *Id.* at *34.

MEALEY'S LITIGATION REPORT: Discovery

- | | |
|--|---|
| 13. <i>Id.</i> at 40. | 19. 2012 WL 4498465 (N.D. Ill. Sept. 28, 2012). |
| 14. 2012 WL 2878130 (S.D.N.Y. July 13, 2012). | 20. <i>Id.</i> at *5. |
| 15. <i>Id.</i> at *12. | 21. <i>Id.</i> |
| 16. No. 6:11-md-2299 (W.D. La. July 27, 2012) (case management order). | 22. <i>Id.</i> at *4. |
| 17. No. CL 61040 (Va. Cir. Ct., Loudoun County, Apr. 23, 2012). | 23. <i>Id.</i> |
| 18. No. 7409-VCL (Del. Ch. Oct. 19, 2012). | 24. <i>Id.</i> |
| | 25. <i>Id.</i> at *5. ■ |

MEALEY'S LITIGATION REPORT: DISCOVERY

edited by Shane Dilworth

The Report is produced monthly by



LexisNexis®

1600 John F. Kennedy Blvd., Suite 1655, Philadelphia, PA 19103, USA
Telephone: (215)564-1788 1-800-MEALEYS (1-800-632-5397)

Email: mealeyinfo@lexisnexis.com

Web site: <http://www.lexisnexis.com/mealeys>

ISSN 1548-6834

EXHIBIT B



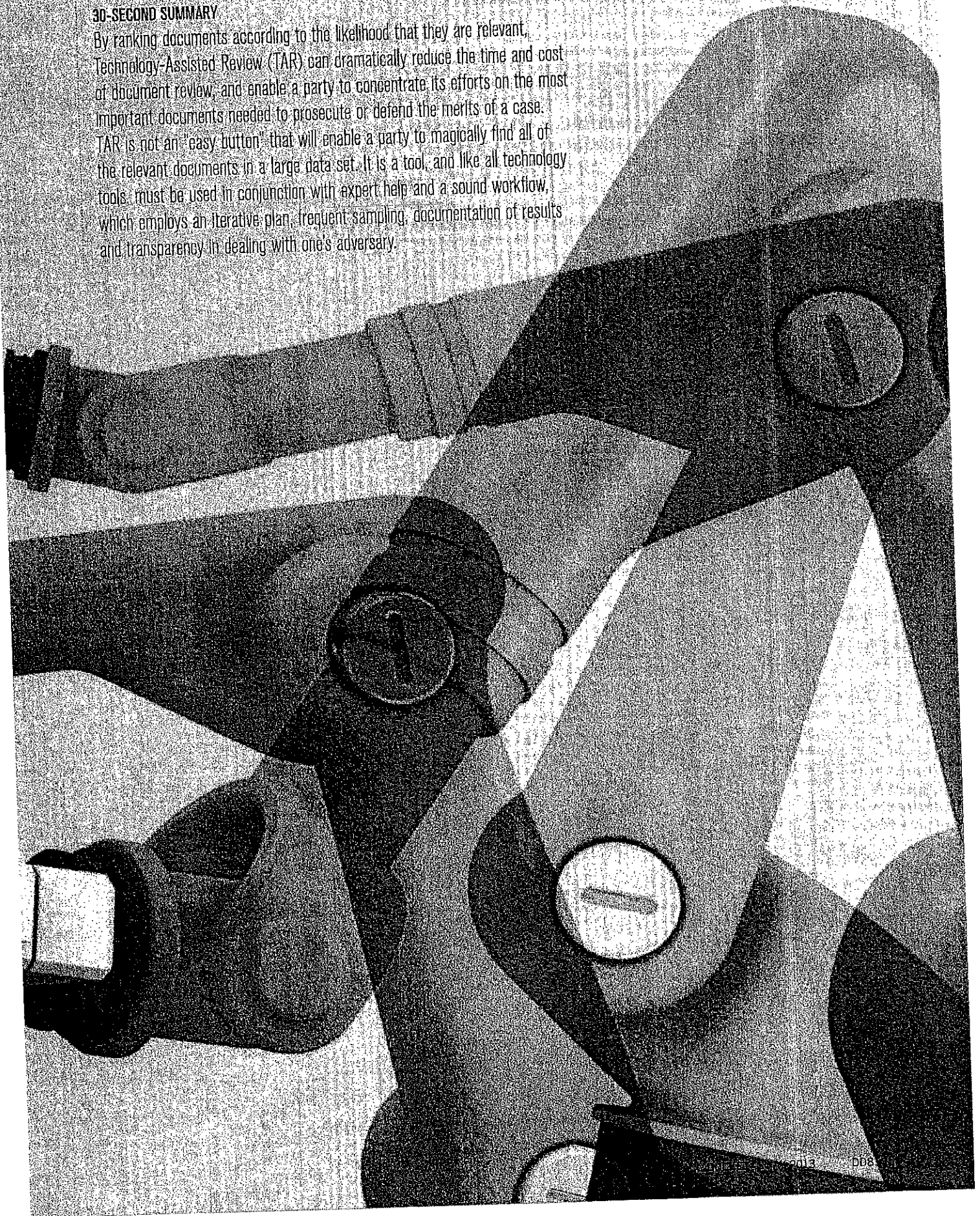
TECHNOLOGY-ASSISTED REVIEW: USE IT OR LOSE IT?

By Cynthia Courtney and Tom Groom

Since the dawn of digital storage, litigation lawyers have conducted document reviews of electronically stored information (ESI) — as they did with paper — one document at a time, or in “linear” fashion. Later, parties used search terms to identify responsive documents, with mixed success at best. Recently, sophisticated prioritization algorithms packaged as ediscovery software applications have exploded onto the scene as a much more efficient and cost-effective identification method, especially in cases involving enormous quantities of ESI. Whether we call these software programs “predictive coding,” “technology-assisted review” or something else entirely, how do you know that they are right for you as a cautious and risk-averse in-house lawyer? Is the science understandable and explainable? Is the cost worth the reward? Is the use of the technology defensible? How do the courts view it? This article will address all of these questions and more.

30-SECOND SUMMARY

By ranking documents according to the likelihood that they are relevant, Technology-Assisted Review (TAR) can dramatically reduce the time and cost of document review, and enable a party to concentrate its efforts on the most important documents needed to prosecute or defend the merits of a case. TAR is not an "easy button" that will enable a party to magically find all of the relevant documents in a large data set. It is a tool, and like all technology tools, must be used in conjunction with expert help and a sound workflow, which employs an iterative plan, frequent sampling, documentation of results and transparency in dealing with one's adversary.



A basic understanding of the science behind technology-assisted review, or TAR, is an important starting point. A number of algorithms and machine-learning technologies underlie leading products in today's ediscovery market. The two primary approaches are concept-based TAR and support vector-based TAR. Both approaches work and both are better than linear or keyword-only workflows. The money spent on TAR is more than offset by substantial savings in review costs, and perhaps most important, "defensibility" flows from the process that is wrapped around the technology and not from the technology itself. Lawyers, not statisticians, are responsible for the completeness of a document production. By constructing workflows and processes that test results at each step of the way and promote iterative improvements, lawyers can confidently argue to an adversary or judge that their results are reasonable and fully defensible.

Basics of TAR statistical models

The ediscovery marketplace uses a number of terms, sometimes interchangeably: predictive coding, computer-assisted review, technology-assisted review. They are terms that ediscovery software developers have created to describe their technology. As noted, we use the term "Technology Assisted Review" (TAR) in this article to describe all of these approaches.

By way of background, computerized technology has greatly affected the efficiencies of document review over the years. Keyword searching — in which a pre-determined set of words are run against all documents, and only those documents that "hit" on one or more key words are reviewed — was a breakthrough technique. Yet, while studies (including the Blair-Marion study and others) conclude that keyword searching only recalls between 20 and 40 percent of relevant documents (see sidebar for

"recall"), keyword search remains the most common approach used today for reducing the number of documents to be reviewed. Keyword search was advanced through methods such as Boolean searches, which allowed the combination of a set of "but not" terms to disambiguate over-inclusive keywords. For example, a search string would state: "Include (keyword1, keyword2, keyword3, etc.) but not (excludeword1, excludeword2, excludeword3, etc.)." These combinations or "ontologies" often included proximity limiters (i.e., Abraham w/2 Lincoln). Studies have shown that ontologies can improve retrieval recall to a range of 65 to 80 percent. However, given the syntax rules that govern ontologies, counsel had to engage highly skilled linguists to develop productive combinations. This was expensive and time-consuming. TAR changes all that by automating this labor-intensive approach.

The primary benefit of TAR is to prioritize documents within a corpus by likelihood of relevance as early as possible in the ediscovery lifecycle. TAR can also be used to prioritize review of client documents or incoming productions and supply a quality check (QC) validation, among other things. In addition to document prioritization, TAR can be used for Early Case Assessment, search-term determination or validation, and ultimately — and perhaps most important — as a means to reduce the amount of data going into a review platform.

The science that underlies TAR is not new; it has been used in other industries, such as energy distribution, air traffic control, weather forecasting and insurance coverage, for decades. Any field where known facts can be extrapolated and monitored with a statistically sound control model can successfully implement this science, known as "predictive analytics." A familiar usage of predictive analytics is calculating credit scores based on previous purchases, buying habits and payment history. Thus, although we hear lawyers express concern about TAR based on a fear that the underlying science is a "black box" that may not work or achieve accurate results, the underlying algorithms are well established and rock solid.

"Concept-based" and "support vector-based" TAR

Although lawyers need not be trained statisticians who fully understand the theoretical minutia behind TAR applications, it is important and helpful to understand certain broad categories of statistical TAR methodology. In both methods discussed below, an "expert" (or "expert team") with a strong knowledge of the facts and claims in the case, known as a subject-matter expert or SME, reviews an initial set of documents and makes a yes/no determination, depending upon whether the document is relevant to the case. (Unless the overall corpus of documents has been filtered, for example, by date, it is possible that a document



Cynthia Courtney, vice president, Discovery Engineering and general counsel, D4 LLC, has more than 10 years of ediscovery experience as both an in-house lawyer at a large health care company and as ediscovery counsel for a law firm. She has written widely and presented at major ediscovery conferences. ccourtney@d4discovery.com



Tom Groom, vice president & senior discovery engineer, D4 LLC, has more than two decades of experience in information technology focused on litigation support, document review and ediscovery methodologies. He advises on issues involving defensible collection methodologies, optimized ESI processing and review workflows, and litigation contingency and readiness planning. tgroom@d4discovery.com

Key statistical terms

- **Confidence level** is expressed as a percentage and represents how often the true percentage lies within the confidence interval. The 95-percent confidence level means that if the test were run 100 times, the same results would be delivered 95 times.
- **Confidence interval** represents the variation from the confidence level for the population. For example, if a confidence interval of ± 2 percent were used with a 95-percent confidence level, the margin of error would be between 93 percent ($95 - 2$ percent) and 97 percent ($95 + 2$ percent). A narrower confidence interval will require a larger sample size. For TAR, a confidence interval of ± 2 percent is common.
- **Recall** asks, "What percent of the relevant documents were retrieved by the algorithm?"
- **Precision** asks, "What percent of a given subset of documents are relevant?" and is inversely proportional to recall.
- **F-measure** is the harmonic mean of recall and precision. The F-measure accounts for the balance between precision and recall where an F-measure reaches its best value at 1 and worst score at 0.

can be *relevant* but outside the date range and, thus, not *responsive*. Therefore, we use the term "relevant" as opposed to "responsive.") Additional review is performed until the system "stabilizes" (meaning the system has learned all it can in order to accurately predict relevance, and no improvement can be made with additional training from the human SME). The entire data set is then run through the algorithm, which scores or ranks the documents. At this point, lawyers and their ESI experts use the ranking information to make decisions about the documents and their review.

Concept-based TAR

Concept-based TAR systems translate the meaning of words used in context within a set of documents into mathematical models based on a conceptual index. Once the model has been built for a document set, a "find more like these" algorithm finds documents that are similar in conceptual content. TAR applications using conceptual search engines require first that statistical parameters be established by the legal team (lawyers and experts) — the confidence level and confidence

interval desired (see sidebar) — followed by the random selection of documents for humans to review and determine content relevance. The subject matter expert (SME) reviews the documents that have been randomly selected by the application. Once the SME has reviewed the initial batch of documents, the SME or an SME team validates the results by sampling documents that have been categorized, and either overturning the relevance decision or allowing the algorithm to rerun until the desired confidence level is met.

Concept-based search engines are a clear improvement over search terms alone. These systems return more potentially content-relevant documents without the limitations of Boolean logic, and the false positives can be suppressed through document seeding (hand-picking exemplar documents that are known to be relevant). The ranking that is obtained from the categorization process measures how closely the documents resemble the exemplars provided to the system. If you are trying to find documents that closely resemble each other, a concept-based TAR application may be for you.

On the plus side, concept-based TAR applications are normally embedded within review platforms. Thus, these applications can be deployed easily if a new set of documents arrives during the review and the legal team decides it may be time to use TAR. Conversely, early use of the technology outside of the review platform is nearly impossible. In addition, some experts criticize concept-based TAR because it tends to take longer to reach stability — and thus to begin the actual review — than SVM-based TAR (described below). This is largely because concept-based TAR requires a post-categorization "overturn correction workflow" to tune the conceptual analysis to the desired confidence level. The SMEs must sample the scored documents, overturn those that they determine have been incorrectly categorized and feed the corrections back into the system for another round of scoring. System stability for concept-based TAR is normally reached after the review of 10,000 to 20,000 documents (8 to 15 days of training). Finally, with most concept-based TAR systems, not all documents are scored. Documents that don't fit into either the "relevant" or "not relevant" category are simply left as "uncategorized."

Support vector machine TAR

The second TAR approach we address is the Support Vector Machine (SVM) methodology. SVM is also a well established "predictive analytics" statistical

The primary benefit of TAR is to prioritize documents within a corpus by likelihood of relevance as early as possible in the ediscovery lifecycle.

modeling and classification methodology, and is used in a variety of industrial applications, including speech recognition, facial expression categorization, handwriting recognition and computational biology. The SVM approach automatically establishes the recall and precision of a given population early in the process (see sidebar). Once recall and precision are determined, the legal team can make informed decisions on the best approach for the prioritized document review.

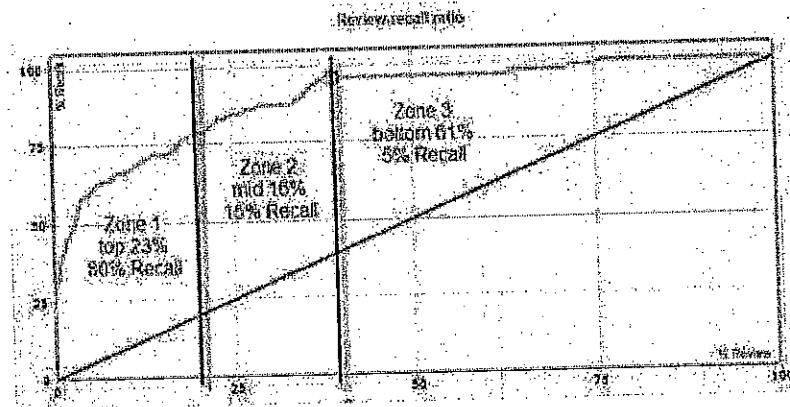
The system first presents a randomly selected set of documents to the SME for review and calculates an initial score. Then the SME reviews several smaller sets of documents that are selected by the system to optimize its learning. During this latter phase, the algorithm includes some documents that have been previously presented to the SME to ensure consistency, as well as documents that are similar to previously selected documents. Behind the scenes, the SVM system is generating a set of weighted attributes to *include* and a subset of weighted attributes to *exclude*, which will result in optimal recall and precision outcomes. At a certain point, using indicators supplied by the TAR engine, the system is said to have “stabilized.” It will then apply a “relevance score” across the entire document population, resulting in a ranking of all documents from the most relevant to the least relevant.

Proponents of the SVM-based approach to TAR appreciate being able to use it outside of the review platform to limit the volume of ESI that must be loaded for review. One clear advantage is that SVM systems typically take only three to five days to train, with a review of 2,500 to 3,500 documents.

Use cases and workflows

TAR applications using either the concept-based or support vector-based algorithm can be used with strong success at several stages of a case. The specific workflow chosen depends

Review-recall ratio



on case variables and the legal team's objectives. However, we believe these workflows, if appropriately documented, strengthen the results and, ultimately, the “defensibility” of a party's production.

1. *Early Case Assessment:* It is early in the case, and counsel anticipates the identification and collection of a large amount of ESI to find documents relevant to the case. TAR can be used on a set of ESI collected from a few key custodians to:

- Find important documents early in the process as part of a settlement risk analysis; and
- Identify an initial set of keywords to use in negotiations with opposing counsel if it is expected that the opposing party will want to use keywords.

2. *Accelerated review:* Counsel faces a very tight time frame with a very large document set. TAR — in particular, the SVM approach — can be used to rank all documents in order of likelihood of relevance; the ranking can be used to make decisions about review priority. Using the relevance scores (see above graph), a review team was able to divide the collection into three zones:

- Zone 1 consists of the top scores for 23 percent of the total document population. Based on the system, Zone 1 contains 80 percent of the relevant documents. From counsel's point of view, this is a “must review” zone. Based on the presumption that every document produced will undergo an “eyes on” review, counsel will conduct a traditional linear review of all documents in this zone.
- Zone 2 consists of the middle 16 percent of the population and contains 15 percent of the relevant documents. Counsel will generate a random sample from this zone to confirm the recall rate and may decide that it need not conduct a linear review of all documents. Counsel will document these results to share during a meet-and-confer session with the opposing party. In the interests of cooperation, counsel will consider offering to share the non-privileged sampled documents themselves in order to substantiate their position. Finally, counsel can come to the meet-and-confer session armed with metrics regarding

the incremental costs of conducting a full review of the documents in Zone 2; with a view to advocating for cost-shifting in the event the opposing party seeks production from that corpus.

- Zone 3 consists of the bottom 61 percent of the document population, but only yields 5 percent of the relevant documents. Counsel will generate a random sample of documents in Zone 3. Upon review of the sample documents, counsel expects to find few relevant documents. They will document their results and likely conclude that further review of the documents in this zone meets no criteria for reasonableness or proportionality of effort for a full initial review.

The firm is thus able to focus its initial review effort on just 23 percent of the population, and in so doing, they expect to identify and produce 80 percent of the relevant documents.

3. *Review QC/verification:* When counsel has completed the review, they turn to quality assurance. They set up a discrepancy matrix, comparing the relevance designations of the review team to the system's relevance scores without reference to the zones described above.

The table (TAR — A reasonable effort) shows there were 3,048 documents that the review team and the TAR application agreed were responsive. There were 40,495 documents that the review team and the TAR application agreed were not responsive.

Of particular interest are the 2,531 documents that the review team marks as not responsive, but that the TAR program scores as responsive. These

documents represent potentially responsive documents that the reviewers may have missed. These 2,531 documents will be submitted to a senior reviewer for second pass review and verification. The senior reviewer finds that almost 1,500 of these documents were in fact responsive. The responsive set increased from 4,624 to 6,000, for an additional one-third on top of the original set that had been slated for production. The lead members of the legal team can comfortably conclude that a reasonable effort has been made to identify responsive documents.

What are the judges and case law saying?

The cases that address TAR are typically — but not always — federal cases decided by United States magistrate judges in the course of managing a matter's discovery. The backstory may include a contentious relationship between opposing counsel. A party may be seeking to use TAR to manage its own immense amount of ESI, or may seek to require its adversary to use TAR to remove as much subjectivity as possible from production decisions. What we glean from the cases, highlighted below, is not an endorsement of TAR itself or a particular algorithm, but guidance about the process and workflow that must accompany the use of TAR to guarantee reliable productions. As the discussion below demonstrates, judges tend to use the term "predictive coding" to refer to any and all TAR applications.

The key questions addressed by the cases are: Will courts *allow* parties to use TAR to assist them in identifying

Additional resources

Maura R. Grossman & Gordon V. Cormack, "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient than Exhaustive Manual Review," *Rich. J.L. & Tech.*, Spring 2011, available at <http://jolt.richmond.edu/v17i3/article11.pdf>

Herbert L. Roitblatt, Anne Kershaw & Patrick Oot, "Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review," 61 *J. Am. Soc'y for Info. Sci. & Tech.* 70, 79 (2010)

David L. Blair & M.E. Maron, "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System," 28 *Comm. ACM* 289 (1985)

production-worthy documents and ESI? Will the court *require* the parties to use TAR? And, where the parties have agreed to use TAR, what are the rules of engagement?

Implicit in the first question is an issue that the courts have *not* addressed — namely, whether a party must disclose its intention to use a TAR application. Traditionally, the method used by a party to identify and collect potentially relevant documents is a matter for determination by that party alone without consultation with the adversary, and in fact, the determination is, at least arguably, subject to the attorney work product privilege, and therefore, shielded from disclosure. Examination of the review methodology would come under scrutiny only if the receiving party suspected that the production was insufficient or inadequate (e.g., if documents relating to a key custodian were absent from the production, or, in the case of ESI, if emails that would logically have

TAR — A reasonable effort

	Review/ Responsive	Review/Not Responsive	
TAR algorithm/responsive	3,048	2,531	5,579
TAR algorithm/not responsive	1,576	40,495	42,071
	4,624	43,026	47,650

TAR applications using either the concept-based or support vector-based algorithm can be used with strong success at several stages of a case.

ACC EXTRAS ON... Ediscovery

ACC Docket

The Five Pillars of In-house Ediscovery (Dec. 2012). www.acc.com/docket/ediscovery_dec12

QuickCounsels

Advanced Technology Solutions for Litigation (Sep. 2010). www.acc.com/quickcoun/atsfl_sep10

Cloud Computing in Ediscovery and Information Governance (Jan. 2012). www.acc.com/quickcoun/info-governance_jan12

ACC HAS MORE MATERIAL ON THIS SUBJECT ON OUR WEBSITE. VISIT WWW.ACC.COM, WHERE YOU CAN BROWSE OUR RESOURCES BY PRACTICE AREA OR SEARCH BY KEYWORD.

responded to messages were not in the production).

It may be that some litigants are using TAR without disclosing that fact to the opposing party. It appears, however, that the more common practice is for a litigant to at least inform the opposing party of its intent to use TAR. This is also the better practice, given the emphasis on transparency and cooperation by the courts when considering the use of TAR.

The first order *allowing* the use of a TAR application over the receiving party's objection was entered on April 23, 2012, in *Global Aerospace, Inc. v. Landow Aviation, L.P., et al.*,¹ a complex dispute in Virginia state court. The defendants sought to use TAR. The plaintiffs opposed the request, arguing that approval of TAR constituted the court's agreement that defendants need only produce the 75 percent of responsive documents the TAR system identifies. Defendants' motion seeking the order was granted. The order is short, stating:

Having heard argument with regard to the Motion of Landow Aviation Limited Partnership, Landow Aviation I, Inc. and Landow & Company Builders, Inc., pursuant to Virginia Rules ... it is hereby ordered Defendants shall be allowed to proceed with the use of predictive coding for purposes of the processing and production of electronically stored information with processing to be completed within 60 days and production to follow as soon as practicable and in no event more than 60 days. This is without prejudice to a receiving party raising with the Court an issue as to completeness or the contents of the production or the ongoing use of predictive coding. (Emphasis is the Court's.)²

At the other end of the spectrum, it appears the show cause order issued in *EORHB, Inc. v. HOA Holdings*³ is the first order *compelling* the use of TAR. In that case, the lower court judge issued a sua sponte order requiring

the parties to use TAR or demonstrate why they should not be required to use it. The Court stated:

This seems to me to be an ideal non-expedited case in which the parties would benefit from using predictive coding. I would like you all, if you do not want to use predictive coding, to show cause why this is not a case where predictive coding is the way to go.

I would like you all to talk about a single discovery provider that could be used to warehouse both sides' documents to be your single vendor. Pick one of these wonderful discovery super powers that is able to maintain the integrity of both sides' documents and ensure that no one can access the other side's information. If you cannot agree on a suitable discovery vendor, you can submit names to me and I will pick one for you.⁴

The *EORHB* story is not over. It is possible the parties will acquiesce to the court's order and agree on a TAR vendor. Or, they may agree that neither wants to use TAR and present that united front in response to the show cause order, supported with arguments about the cost of TAR. In that event, the judge may want to consider hiring his own expert to advise him on the real costs of TAR — i.e., the cost of TAR when balanced against potential review savings.

Can a court *compel* a party to use TAR over its objection? Not so far. In *Kleen Products LLC, et al. v. Packaging Corp. of America*,⁵ after a protracted struggle in which plaintiffs in an antitrust case sought to require defendants to use TAR, the plaintiffs withdrew their demand, but only as to the document request already on the table, to which defendants had responded, and pursuant to which documents had been produced following the retention of an ediscovery vendor and the application of search terms. The parties argued their respective positions in two separate hearings and a substantial number of written submissions. For future production

HAVE A COMMENT ON THIS ARTICLE? VISIT ACC'S BLOG AT WWW.INHOUSEACCESS.COM/ACC-DOCKET.

requests, the parties agreed to meet and confer regarding production methods, raising the possibility that plaintiffs will renew their demands for the use of TAR.

The best-known case addressing TAR is *DaSilva Moore v. Publicis Groupe and MSL Group*,⁶ which commanded considerable attention in the ediscovery community in early 2012. In that case, the issue appears to have been the manner in which TAR would be employed. In his Feb. 24, 2012,⁷ opinion, US Magistrate Judge Andrew Peck went to great pains to assure the parties that he did not believe TAR is appropriate for all cases or even all large cases; further, that parties intending to use TAR must choose a reliable vendor and program; and perhaps most important, must design an "appropriate process" that includes "appropriate quality control testing."

US District Judge Andrew Carter, in reviewing Judge Peck's decision along with plaintiff's request that Judge Peck recuse himself, stated that the protocol approved by Judge Peck "contains standards for measuring the reliability of the process and the protocol builds in levels of participation by Plaintiffs. ... That the search methods will be carefully crafted and tested for quality assurance, with Plaintiffs participating in their implementation."⁸ For example, plaintiffs' counsel were permitted to submit keywords and review the resulting documents before the documents were produced, retaining the ability to object to the protocol at that time. Plaintiffs were also given the ability to object once production was made if they believed it to be insufficient. Judge Carter concluded that to call the proposed TAR method unreliable at that stage of the proceedings would be speculative.⁹

Judge Peck's opinion lays out a number of steps the parties should take to assure the reliability of a party's use of TAR ("predictive coding"):

1. Bring both vendor experts to a court hearing to respond to the magistrate judge's questions;

2. Allow the requesting party to view the documents that were used to train the TAR system, both those that were marked responsive and those that were marked non-responsive;
3. Allow the requesting party to view the additional documents that were used to "stabilize" the TAR system, whether or not marked responsive;
4. Do not adhere to an arbitrary number of documents that will be produced, without reference to the statistical results; and
5. Do not limit up front the number of iterative reviews used to "train" the system, but rather, assessing whether the system had "stabilized" before stopping the iterative reviews.¹⁰

The principles of cooperation, transparency and attention to a strong workflow that are emphasized in *DaSilva Moore* have a concrete and defined role in *In re Actos Products Liability Litigation*,¹¹ and in fact, are taken a step further. In the *In re Actos* case, parties entered into a detailed protocol, wherein three experts from each side were selected to meet, collaboratively review the training set of defendants' documents together, and agree upon relevance determinations for all non-privileged documents in the set. Robust protections to guard documents subject to privilege and confidentiality are included. The *In re Actos* protocol is a great resource for parties considering the use of TAR.

A new tool for ESI discovery

Technology-assisted review (TAR) is a provocative, exciting and, for some, intimidating litigation support technology. TAR has emerged at a time when vast quantities of ESI threaten to overwhelm litigants and the American court system. By organizing documents according to the likelihood that they are relevant, TAR can dramatically

reduce the time and cost of document review and enable a party to concentrate its efforts on the most important documents needed to prosecute or defend the merits of a case.

TAR is not an "easy button" that will magically find all of the relevant documents in a large data set. It is a tool, that must be used in conjunction with expert help and a sound workflow, employing an iterative plan, frequent sampling, documentation of results and transparency in dealing with your adversary. **ACC**

NOTES

- 1 *Global Aerospace Inc., et al. v. Landow Aviation L.P. d/b/a Dulles Jet Center et al.*, No. CL 61040 (Cir. Ct. Loudon Cnty., April 23, 2012) (order allowing Defendant to proceed with predictive coding)
- 2 *Id.*
- 3 Transcript of Motion for Partial Summary Judgment, Motion to Dismiss Counterclaim and Ruling of the Court at 66-67, *EORHB, Inc. v. HOA Holdings, LLC*, No. 7409-VCL (Del. Ch. Oct. 15, 2012).
- 4 *Id.*
- 5 *Kleen Products LLC, et al. v. Packaging Corp. of America, et al.*, Civ. No. 1:10-cv-05711 (Aug. 21, 2012 N.D. IL)
- 6 *DaSilva Moore v. Publicis Groupe and MSL Group*, 11 Civ. 1279 (ALC) (AJP) (SDNY)
- 7 *Id.* (2/24/12)
- 8 *DaSilva Moore, et al. v. Publicis Groupe, et al.* 11 Civ. 1279 (ALC) (AJP) (11/7/12)
- 9 *Id.*
- 10 *Id.* (2/24/12)
- 11 *In re Actos Products Liability Litigation*, MDL No. 6:11-md-2299 (July 27, 2012)

Reprinted with permission from the Association of Corporate Counsel, © 2013. All Rights Reserved. www.acc.com

EXHIBIT C

RECEIVED

JUL 27 2012 *JO*TONY R. MOORE, CLERK
WESTERN DISTRICT OF LOUISIANA
LAFAYETTE, LOUISIANAUNITED STATES DISTRICT COURT
WESTERN DISTRICT OF LOUISIANAIn Re: Actos (Pioglitazone) Products
Liability Litigation

MDL NO. 6:11-md-2299

JUDGE DOHERTY

This Document Applies to:

MAGISTRATE JUDGE HANNA

All Cases

CASE MANAGEMENT ORDER:
PROTOCOL RELATING TO THE PRODUCTION OF
ELECTRONICALLY STORED INFORMATION ("ESI")

Pursuant to the agreement reached between the Plaintiffs and Defendants herein, this Court enters the following Order concerning the production of electronically stored information in these proceedings:

A. Scope

1. General. The procedures and protocols outlined herein govern the production of electronically stored information ("ESI") by the Parties. Section E titled "Search Methodology Proof of Concept" applies only to the predictive coding and advanced analytics sampling procedure as outlined in that Section. Sections A through D and Sections F through J apply throughout the pendency of this litigation. This Order governs all parties to these proceedings, whether they currently are involved or become so in the future. The Parties to this protocol ("Protocol") will take reasonable steps to comply with this agreed-upon Protocol for the production of documents and information existing in electronic format. All disclosures and

productions made pursuant to this Protocol are subject to the Privilege Protocol and Protective Order entered in this matter.

2. Limitations and No-Waiver. The Parties and their attorneys do not intend by this Protocol to waive their rights to the attorney work-product privilege, except as specifically required herein, and any such waiver shall be strictly and narrowly construed and shall not extend to other matters or information not specifically described herein. All Parties preserve their attorney client privileges and other privileges and there is no intent by the protocol, or the production of documents pursuant to the protocol, to in any way waive or weaken these privileges. All documents produced hereunder are fully protected and covered by the Parties' confidentiality agreements, and order(s) of the United States District Court, as well as any clawback agreements, and protective order(s) of the United States District Court effectuating same.

B. ESI Preservation

1. The Parties have issued litigation notices to those identified as most likely to have discoverable information.

C. Sources

1. While Defendants' fact gathering is ongoing, the following are data sources identified to date that are most likely to contain discoverable information. Defendants agree to provide additional discovered data sources likely to contain relevant information. Defendants agree to provide information about the data sources to the extent applicable and known in addition to that found in the subparagraphs below, including the date range of information contained in the data source, the department(s) utilizing the data source, whether the data source is hosted internally or externally, and the database type.

a	ARISg	Adverse Event Database
b	BLUE	Labeling and promotional materials management system
c	Galaxy	Regulatory document management system
d	MEDIsorce	Product information request database
e	T-Rx	Field sales call database
f	TSARS (or "S Drive")	Takeda Statistical Analysis and Repository System
g	T-Track	Clinical Science Liaison database
h	IRIS	Research grant management system
i	LARC	Clinical Science Liaison education resources database
j	Sample Guardian	Product sample management database
k	TEG	Takeda Educational Grant management system
l	PubBase	Publications management system
m	Records Management System	Records Operation Center ("ROC") information system

a. ARISg: ARISg is an adverse event database. It contains information that the Pharmacovigilance department at TRGD U.S. receives regarding adverse events related to Takeda drugs, including adverse event reports ("AERs") received from, without limitation, physicians, patients, clinical trials, medical literature, and foreign entities. ARISg is the software used for this database, which is sometimes called T-Gaea within Takeda. It has been in effect since 1999.

b. BLUE: This database is used by the Marketing department in the approval process for promotional materials. It contains a labeling module and a module for promotional pieces and marketing campaigns. BLUE has been active from April 2008 to present. The vendor is Schawk Blue.

c. Galaxy: Galaxy is a document repository system used by the Regulatory department containing components of regulatory submissions to the Food and Drug Administration. It went into production in 2009.

d. MEDIsorce: This data system is used by the Medical Information and Quality Assurance departments to capture and respond to product information requests and non-medical product complaints. It has a Siebel component that documents the intake of requests for information from physicians and provides a response; a Documentum system with standard response and customer response letters; and Info Maestro which pulls information from the standard response letter and from the Siebel system to create the response letter to an individual physician.

e. T-Rx: This database contains information regarding U.S. commercial field sales calls.

f. TSARS (or "S Drive"): This is Takeda's Statistical Analysis and Repository System and is a Unix centralized repository used to manage Clinical and research data. It is used by the Analytical Science department. It contains clinical SAS data sets and programs used to analyze those data sets for purposes of final submission reports – tables, listings, and graphs.

g. T-Track: This database is a customized application of Seibel's Customer Relationship Management system for use by Takeda's field based Clinical Science Liaisons.

h. IRIS: This system is used by Takeda for the intake and processing of external research grant requests. It is a vendor hosted system (SteepRock is the vendor). It was implemented within the last five years.

i. LARC: This database includes articles, presentations, and publications related to Takeda products and the therapeutic areas they address. Quosa is the vendor for this database. It is accessible by Clinical Science Liaisons in their respective therapeutic areas.

j. Sample Guardian: This database contains product sample management data regarding sample transactions and inventory reconciliations.

k. TEG: Takeda Educational Grant database is used for education grant request management.

l. PubBase: PubBase is a Documentum-based system used for the management and storage of publication documents.

m. Records Management System: This data source is used by the Records Operations Center ("ROC"), where physical records are maintained.

D. Custodians

1. The following are custodians who have been identified as most likely to have information relevant to this litigation. For these custodians, data is being pulled from e-mail, computer hard drives, and physical files that are in the possession, custody, and control of Takeda. Investigation is ongoing by both Parties as to potential additional custodians at Takeda (including potential Japanese custodians) and Eli Lilly and Company. Current key custodians include:

1.	Baron, David	Vice President, NonClinical Safety/Efficacy
2.	Spanheimer, Robert	Vice President, Medical and Scientific Affairs
3.	Greeby, Jennifer	Director, Marketing (Diabetes)
4.	Recker, David	Senior Vice President, Clinical Science
5.	Paris, Maria	Former Vice President, Pharmacovigilance
6.	Gerrits, Charles	Former Senior Director, Pharmacoeepidemiology
7.	Johnston, Janet	Associate Director, Safety Surveillance
8.	Thom, Claire	Former Vice President, Research and Development
9.	Daly, Rich	Former Vice President, Marketing
10.	Perez, Alfonso	Vice President, Clinical Science Strategy

11.	Ortell, Una	Director, Promotion and Advertising
12.	Orlando, Dan	Former Vice President, Sales
13.	Lee, Jessie	Manager, Regulatory Affairs Strategy
14.	Cuomo, Maryann	Associate Director, Regulatory Labeling
15.	Weisbrich, Shay	Vice President, Franchise Leader (Former Director, Marketing)
16.	Kupfer, Stuart	Vice President, Clinical Science
17.	Ramstack, Mary	Sr. Director, Strategic Project Planning and Management
18.	Roebel, Mick	Sr. Director, Regulatory Affairs
19.	Lorenz, Janet	Associate Director, Regulatory Affairs, Promotion and Advertising
20.	Pritza, Mary Jo	Former Associate Director, Regulatory Affairs
21.	Caracci, Mike	Former Director, Marketing
22.	Tynan, Julie	Assistant Project Director, Strategic Project and Planning Management
23.	Hull, Andy	Vice President, Alliance Management (former Vice President, Marketing)
24.	Fusco, Gregory	Sr. Medical Director, Pharmacoeconomics and Analysis
25.	Caggiano, Christopher	Sr. Product Manager, Diabetes Marketing
26.	Ryan, D'Arcy	Former Director, Marketing
27.	Khan, Mehmood	Former Sr. Vice President, Medical and Scientific Affairs
28.	Harris, Thomas	Vice President, Regulatory Affairs
29.	Trochanov, Anton	Associate Medical Director, Pharmacovigilance

E. Search Methodology Proof of Concept

1. General. The Parties have discussed the methodologies or protocols for the search and review of ESI collected from Takeda sources, including but not limited to e-mail, and the following is a summary of the Parties' agreement on the use of a search methodology proof of concept to evaluate the potential utility of advanced analytics as a document identification mechanism for the review and production of this data. The Parties agree to meet and confer regarding the use of advanced analytics for other data sources. While the Parties agree to explore the use of advanced analytics as a technique to ensure appropriate responses to discovery

requests, the Parties agree that Defendants retain the right to review documents after predictive coding but prior to production for relevance, confidentiality, and privilege. A sampling of documents withheld after such review will take place pursuant to Section E.10.

2. General Overview of Advanced Analytics/Predictive Coding Process. Takeda utilizes software provided by Epiq Systems (“Epiq”) to search and review ESI for production in this case. Epiq uses Equivio’s Relevance software for advanced analytics and predictive coding.

Epiq will collect e-mail documents from four key Takeda custodians, which will be combined to create the “sample collection population.” The Parties will meet and confer to determine the names of the four custodians. Additionally, Takeda will add a set of regulatory documents which have already been collected to the “sample collection population.” Takeda and Plaintiffs will each nominate three individuals (“the experts”) to work collaboratively at the offices of Nelson Mullins, 1320 Main Street, Columbia, SC 29201 to train the Equivio Relevance system. Plaintiffs’ experts will execute a Nondisclosure and Confidentiality Agreement in the form attached as Exhibit A hereto. To the extent that Plaintiffs’ experts are exposed to information that would be subject to withholding or redaction under the Protective Order in this matter, Plaintiffs’ experts agree not to disclose such information to co-counsel, client, any Party, or any third party without obtaining prior written consent of the other Party regarding the particular piece of information sought to be disclosed. Before the meeting, the Parties shall be provided a copy of the applicable Equivio training documents, handbook, or manual. The Parties’ experts will receive technical training on the Equivio Relevance software and coding process and will work together to make one relevance decision for documents in the Control and Training sets, as described in more detail below.

The Parties will review a number of documents required by the Equivio Relevance system for the data to reach Stability as described below. Once Stability is reached, the Control and Training sets are then used to begin the predictive coding process. Using the Control and Training documents, the system calculates relevance scores for the entire sample collection population, with each document in the sample collection population receiving a relevance score of 0 through 100.

Attorneys representing Takeda will have access to the entire sample collection population to be searched and will lead the computer training, but they will work collaboratively with Plaintiffs' counsel during the Assessment and Training phases. Takeda's experts will conduct an initial review of documents presented by the Equivio Relevance system for privilege. The privileged documents will be either entirely withheld from viewing by Plaintiffs' experts or printed and redacted. A privilege log for such documents will be provided. The Parties, after review of the privilege log, reserve the right to require that such documents be deemed as "skip" (same as designation used for technical problem documents). Otherwise, these documents may still be used to train the system. Both Parties will then review all of the non-privileged documents during the training process (i.e., both documents coded as relevant and irrelevant). The Parties' experts will review the documents in collaboration and determine the coding to be applied to the documents. To the extent the Parties disagree regarding the coding of a particular document or designation of privilege, they will meet and confer in an effort to resolve the dispute prior to contacting the Court for resolution.

At the conclusion of the training process and upon calculation of relevance scores, the Parties will meet and confer regarding which relevance score will provide a cutoff for documents

to be manually reviewed by defense counsel for production. However, the Parties reserve the right to seek relief from the Court prior to the commencement of the final manual review.

At the recommendation of Epiq, no seeding will take place at this time. The Parties may meet and confer if it is determined that seeding may be applicable at a later date.

Plaintiffs' experts and counsel shall not remove any of the Control or Training documents from the offices of Nelson Mullins, nor shall they be allowed to copy such documents. The Parties agree that Defendants do not waive protection of trade secret or confidential information in allowing Plaintiffs to review documents under this sampling mechanism. All documents reviewed pursuant to this sampling protocol shall be done under the Protective Order in this matter as well as any Privilege Protocol or clawback agreement that shall be reduced to an order acceptable to the Court.

3. Relevance Tags. The Parties agree that as part of the Assessment and Training phases, all of the non-privileged and privilege-redacted documents reviewed by both parties' experts will be categorized as relevant, not relevant, or skip (to be used for documents with technical problems). The privileged-withheld documents will be categorized by Defendants' experts as relevant, not relevant, or skip, subject to the Parties' right to have any privileged-withheld documents categorized as a "skip." The Parties shall immediately discuss any disagreements on coding in good faith, so that the training may be improved accordingly, and may seek guidance from the Court or the Court appointed special masters if necessary.

4. Collection & Data Preparation. The Parties will meet and confer to agree upon the four custodians that will be selected for the sampling. E-mail and attachment documents will be collected from the four custodians and added to the collected regulatory documents, together

comprising the sample collection population. Documents may be removed from the sample collection population if they are:

- a. Spam,
- b. Commercial e-mail,
- c. Files without text,
- d. Exact duplicates within the custodians (see Section G.6 regarding production of information for duplicate documents), and
- e. System files, etc. (*i.e.*, the documents that the samples will be selected from will be de-NISTED)

Epiq will extract the sample collection population documents' text and build an index.

5. Assessment Phase. The Equivio Relevance software generates an initial simple random sample of 500 documents from the sample collection population. Takeda's experts will initially review the documents for privilege. Any documents deemed privileged by Takeda's experts will be either entirely withheld from viewing by Plaintiffs' experts or printed and redacted prior to viewing by Plaintiffs' experts, and logged on a privilege log consistent with the Privilege Protocol in this matter. These documents may still be used to train the system. To the extent the Parties disagree regarding the privilege decision for a particular document, they will meet and confer in an effort to resolve the dispute prior to contacting the Court for resolution. The Parties' experts will then work collaboratively to determine the relevance of the non-privileged and privilege-redacted documents. The relevance of the privileged-withheld documents will be determined by Defendants' experts. The documents reviewed in the Assessment Phase make up the Control Set. The Control Set is used for estimating richness

(percentage of relevant documents in a population), and also serves as a reference point for calculating recall and precision.

a. The application's estimates of richness use a confidence level of 95%. The initial Control Set of 500 documents yields a confidence estimation of richness with an error margin of plus or minus 4.3%. This is a worst-case error margin assuming richness of 50%. For lower levels of richness, the error margin will also be lower. For example, for richness of 10%, the error margin would be plus or minus 2.6%, while for 5%, the error margin would be plus or minus 1.9%.

b. The Control Set also creates a basis for calculating recall and precision, which are then used for monitoring training progress and calculating results.

c. Equivio Relevance tracks the progress of the Assessment Phase to achieve the appropriate level of statistical validation. These levels of validation are referred to in the Equivio system as "Baseline," at the lowest level, through "Statistical," at the highest level. The terms "Baseline" and "Statistical" are used by Equivio Relevance as indicators to the user as to the progress of the Assessment Phase. The validation level achieved depends on the number of relevant documents found by the user in the Control Set. At the "Baseline" level, the number of relevant documents in the control set is too low to allow statistically valid estimates of recall and precision. The Parties will ensure that the number of Control Set documents reviewed will reach the "Statistical" level.

d. For informational purposes, the "Statistical" level of validation in Equivio requires the presence of at least 70 relevant documents in the Control Set. For document collections with richness of 14% and above, a Control Set of 500 documents is sufficient to reach

the “Statistical” level of validation. For lower levels of richness, additional documents will need to be reviewed in the Assessment Phase in order to reach the “Statistical” level.

e. Based on a confidence level of 95%, the Statistical level of validation yields an error margin on recall estimates of plus or minus 11.7%. This is a worst-case error margin assuming recall of 50%. The Parties will continue the Assessment Phase, beyond the “Statistical” level, until the Control Set contains at least 385 relevant documents. This sample will yield an error margin on recall estimates of plus or minus 5%.

6. Iterative Training Phase. Following the creation of the Control Set at the Statistical validation level, the Equivio Relevance system selects a random sample of forty documents. Takeda’s experts will initially review the forty documents for privilege. Any documents deemed privileged by Takeda’s experts will be either entirely withheld from viewing by Plaintiffs’ experts or printed and redacted prior to viewing by Plaintiffs’ experts, and logged on a privilege log consistent with the Privilege Protocol in this matter. These documents may still be used to train the system. The Parties’ experts will then work collaboratively to determine the relevance of the non-privileged and privilege-redacted documents. The relevance of the privileged-withheld documents will be determined by Defendants’ experts, subject to the Parties’ right to have any privileged-withheld documents categorized as a “skip” and not included in the training. To the extent the Parties disagree regarding the relevance or privilege decision for a particular document, they will meet and confer in an effort to resolve the dispute prior to contacting the Court for resolution.

a. Once the experts have completed the first Training Set, the Equivio Relevance system calculates the Training Status. The three possible states are “Not Stable,” “Nearly Stable,” or “Stable.”

b. The experts continue to review samples of forty documents each, using the process outlined in paragraph 6 above, until the Stable Training Status is reached.

c. The subsequent samples of forty documents are selected using an Active Learning approach. Active Learning means that each training sample is selected based on what has been learned from previous samples. The object is to maximize the sample's contribution to the training process. Therefore, the system chooses samples that provide comprehensive coverage of the population (reducing under-inclusiveness), while fine-tuning the concept of relevance that the Classifier is developing (reducing over-inclusiveness). The system reaches Stability when the marginal contribution of additional samples to the enhancement of the Classifier approaches zero, as determined by the Equivio software and which determination (Stability) is not configurable.

7. Calculation of Relevance Scores. Upon completion of the Training Phase once Stability is reached, and any related meet and confer sessions and agreed upon coding corrections, the Equivio Relevance system will run over the sample collection population and calculate relevance scores for each document in the sample collection population. Each document in the sample collection population receives a relevance score of 0 through 100, with 0 being least likely to be relevant and 100 being most likely.

8. Final Search, Review, and Production of Sample Collection Population Documents. The Parties will meet and confer regarding which relevance score will provide a cutoff that will yield a proportionate set of documents that will be manually reviewed by Takeda for production. All of the documents above the agreed upon relevance score in the sample collection population will be reviewed by Takeda. Documents found by Takeda's review to be relevant and non-privileged documents will be produced to Plaintiffs.

9. Quality Control by Random Sample of Irrelevant Documents. In addition, at the conclusion of the process described above, and prior to generating the review set, the Parties will collaboratively review at the offices of Nelson Mullins in Columbia, SC a random sample of documents in the sample collection population with relevance scores below the cut-off score set for establishing the review set (aka the "Rest"). These documents are flagged for culling, and will not be included in the review set. In Equivio Relevance, this test is referred to as "Test the Rest." The purpose for this phase is to verify that the Rest contains a low prevalence of relevant documents and that the proportionality assumptions underlying the cut-off decision are valid.

a. The Test the Rest sample is designed to provide a confidence level of 95%. The default sample size is 500 documents. The margin of error depends on the percentage of relevant documents in the Rest. For example, if 5% of the Rest documents are found to be relevant, the margin of error is 1.9%. If 1% are relevant, the margin of error is 0.8%.

b. Takeda's experts will initially review the Rest sample documents for privilege. Any documents deemed privileged by Takeda's experts will be either entirely withheld from viewing by Plaintiffs' experts or printed and redacted prior to viewing by Plaintiffs' experts, and logged on a privilege log consistent with the Privilege Protocol in this matter. The Parties' experts will then work collaboratively to determine the relevance of the non-privileged and privilege-redacted documents. The relevance of the privileged-withheld documents will be determined by Defendants' experts, subject to the Parties' rights to have any privilege-withheld document categorized as a "skip" for purposes of the Test the Rest sample. To the extent the Parties disagree regarding the relevance or privilege decision for a particular document, they will meet and confer in an effort to resolve the dispute prior to contacting the Court for resolution.

10. Sampling of Documents Not Produced After Predictive Coding. After the predictive coding process completes, and Takeda's counsel reviews and produces documents from the sample collection population consistent with paragraph 8, the Parties will collaboratively review at the offices of Nelson Mullins in Columbia, SC a random sample of documents above the agreed-upon cutoff relevance score that were withheld from production on relevance grounds. The Parties agree to meet and confer regarding an appropriate sample size.

a. Takeda's experts will initially review the sample documents for privilege. Any documents deemed privileged by Takeda's experts will be either entirely withheld from viewing by Plaintiffs' experts or printed and redacted prior to viewing by Plaintiffs' experts, and logged on a privilege log consistent with the Privilege Protocol in this matter. The Parties' experts will then work collaboratively to determine the relevance of the non-privileged and privilege-redacted documents. The relevance of the privileged-withheld documents will be determined by Defendants' experts, subject to the Parties' rights to have any privilege-withheld document categorized as a "skip" for this purpose. To the extent the Parties disagree regarding the relevance or privilege decision for a particular document, they will meet and confer in an effort to resolve the dispute prior to contacting the Court for resolution.

11. Post-Predictive Coding Sampling Meet and Confer. The Parties shall meet and confer in good faith to resolve any difficulties and finalize the method for searching documents on a going forward basis. To the extent that the Parties cannot agree, they shall apply to the Court for relief. Defendant shall not be required to proceed with the final search and review unless and until objections raised by either Party have been adjudicated by the Court or resolved by written agreement of the Parties. The Parties reserve the right to request a meet and confer

regarding the designation of any document as a “skip” for purposes of the control sample, training, or Test the Rest, if agreement cannot be reached.

F. Costs

1. Takeda reserves its right to seek relief from the Court (e.g., a cost shifting award and pursuant to the principles of proportionality). *See* Fed. R. Civ. P. 1, 26(b)(2)(C), 26(b)(2)(B), & 26(g); *Electronic Discovery*, 11 Sedona Conf. J. 289 (2010); *see also* Fed. R. Evid. 403 (inadmissibility of cumulative evidence).

2. Plaintiffs agree to bear all of the costs associated with their compliance with the terms of this protocol. Plaintiffs agree to bear all of the costs associated with the receipt and review of ESI produced hereunder including the costs associated with its ESI experts who will be involved with Plaintiffs in all aspects of this ESI protocol.

G. Format of Production For Documents Produced by Defendants

1. TIFF/Native File Format Production. Documents will be produced as single-page TIFF images with corresponding multi-page text, native file format document if applicable under paragraph G.2, and necessary load files. Native files, along with all corresponding metadata, will be preserved. TIFF images will be of 300 dpi quality or better. The load files will include an image load file as well as a metadata (.DAT) file with the metadata fields identified below on the document level to the extent available.

	Field	Summation Field (Florida)	Definition	Doc Type
1	SOURCE	SOURCE	Name of party producing the document	All
2	CUSTODIAN	CUSTODIAN	Name of person or non-human data source from where documents/files are produced. <i>**Where redundant names occur, individuals should be distinguished by an initial which is kept constant throughout productions (e.g., Smith, John A. and Smith, John B. Where data is collected from an archive, the archive will be listed as custodian.</i>	All
3	CUSTODIANAPPENDMULTI	CUSTODIANAPPENDMULTI	Name of Takeda person or non-human data source from where duplicate documents/files were suppressed. <i>**Where redundant names occur, individuals should be distinguished by an initial which is kept constant throughout productions (e.g., Smith, John A. and Smith, John B. Where data is collected from an archive, the archive will be listed as custodian.</i>	All
4	CUSTODIAN ID	CUSTODIAN ID	Each CUSTODIAN from #2 or 3 above will be assigned a unique numeric identifier that will be maintained throughout productions. Where data is collected from an archive, the archive will be listed as custodian.	All

	Field	Summation (Florida)	Field	Definition	Doc Type
5	BEGBATES	BEGDOC#		Beginning Bates Number (production number)	All
6	ENDBATES	ENDDOC#		End Bates Number (production number)	All
7	PGCOUNT	PGCOUNT		Number of pages in the document	All
8	FILESIZE	FILESIZE		File Size	All
9	APPLICAT	APPLICAT		Commonly associated application for the specified file type.	All
10	FILEPATH	FILEPATH (for Edocs)		File source path for electronically collected documents other than emails, which includes location, file name, and file source extension.	Edocs
11	RELATIVE PATH APPEND	RELATIVE PATH APPEND (for Edocs)		File source path for duplicate electronically collected documents other than emails, which includes location, file name, and file source extension.	Edocs
12	NATIVEFILELINK	DOCLINK		For documents provided in native format only	All
13	TEXTPATH	LOGFILE or FULLTEXT		File path for OCR or Extracted Text files	All
14	MSGID	MSGID		Value extracted from parent message during processing	Email
15	FROM	FROM		Sender	Email
16	TO	TO		Recipient	Email
17	cc	cc		Additional Recipients	Email
18	BCC	BCC		Blind Additional Recipients	Email
19	SUBJECT	SUBJECT		Subject line of email	Email
20	PARENTBATES	PARENTID		BeginBates number for the parent email of a family (will not be populated for documents that are not part of a family)	Email

	Field	Summation (Florida)	Field	Definition	Doc Type
21	ATTACHBATES	ATTACHID		Bates number from the first page of each attachment	Email
22	BEGATTACH	(will be provided from ATTRANGE)		First Bates number of family range (i.e. Bates number of the first page of the parent email)	Email
23	ENDATTACH	(will be provided from ATTRANGE)		Last Bates number of family range (i.e. Bates number of the last page of the last attachment)	Email
24	ATTACHCOUNT	ATTACHMENT COUNT		Number of attachments to an email	Email
25	ATTACHNAME	ATTACHMENT LIST		Name of each individual attachment	Email
26	DATESENT (mm/dd/yyyy hh:mm:ss AM)	DATESENT		Date Sent	Email
27	DATERCVD (mm/dd/yyyy hh:mm:ss AM)	DATERCVD		Date Received	Email
28	EMAILDATSORT (mm/dd/yyyy hh:mm:ss AM)	DATSENT		Sent Date of the parent email (physically top email in a chain, i.e. immediate/direct parent email)	Email
29	Email Outlook Type	Email Outlook Type		Type of Outlook item, e.g. email, calendar item, contact, note, task	Email
30	HASHVALUE	MD5HASH		MD5 Hash Value	All
31	TITLE	DOCTITLE		Title provided by user within the document	Edocs
32	AUTHOR	AUTHOR		Creator of a document	Edocs
33	DATECRTD	DATECRTD		Creation Date	Edocs
34	MODIFIED BY	LAST EDITED BY		Person who has modified a document	Edocs
35	LASTMODD (mm/dd/yyyy	LASTMODD (mm/dd/yyyy hh:mm:ss		Last Modified Date	Edocs

	Field	Summation (Florida)	Field	Definition	Doc Type
36	DocumentType	DocumentType		Descriptor for the type of document: "E-document" for electronic documents not attached to emails; "Emails" for all emails; "E-attachments" for files that were attachments to emails; and "Physicals" for hard copy physical documents that have been scanned and converted to an electronic image.	All
37	Importance	Importance		High Importance - indicates Priority Email message.	Email
38	Redacted	Redacted		Descriptor for documents that have been redacted. "Yes" for redacted documents; "No" for unredacted documents.	All
39	ProdVol	ProdVol		Name of media that data was produced on. Wave 00 I - Hard Drive	All
40	Confidentiality	Confidentiality		Indicates if the document has been designated as "Confidential" pursuant to any applicable Protective Order. "Yes" for Confidential documents; "No" for documents that are not so designated.	All
41	Email folder	Email folder		Folder in which non-archive collected email is stored within the custodians mailbox, such as "inbox", "sent", "deleted", "draft", or any custom folder.	Email

	Field	Summation Field (Florida)	Definition	Doc Type
42	Relevance score	Relevance score	Relevance score assigned by Equivio for documents that have been through the predictive coding process	All

a. This list of fields does not create any obligation to create or manually code fields that are not automatically generated by the processing of the ESI; that do not exist as part of the original Metadata of the document; or that would be burdensome or costly to obtain.

2. Defendants will produce spreadsheets (.xls/.xlsx files) and PowerPoint presentations (.ppt/.pptx files) in native form as well as audio and video files (e.g., mp3s, wavs, mpegs, etc.), except that spreadsheets and PowerPoint documents will be produced in TIFF format if redactions are applied. Audio and video files shall be edited if redactions are required, subject to appropriate identification of any such audio or video files having been edited. In addition, for any redacted documents that are produced, the documents' metadata fields will be redacted where required. The Parties will meet and confer regarding a request for the production of any other materials including documents in native file format.

3. The Parties agree to meet and confer regarding the format of production for structured databases.

4. Appearance. Subject to appropriate redaction, each document's electronic image will convey the same information and image as the original document, including formatting, such as bolding, highlighting, font size, italics. Documents will be produced in black and

white. After production, a Party may request that a document be produced in color at which time the Parties may meet and confer about such production. Documents that present imaging or formatting problems will be identified and the Parties will meet and confer in an attempt to resolve the problems.

5. Document Numbering. Each page of a produced document will have a legible, unique page identifier "Bates Number" electronically "burned" onto the image at a location that does not obliterate, conceal or interfere with any information from the source document. The Bates Number for each page of each document will be created so as to identify the producing Party and the document number. In the case of materials redacted in accordance with applicable law or confidential materials contemplated in any Protective Order or Confidentiality Stipulation entered into by the Parties, a designation may be "burned" onto the document's image at a location that does not obliterate or obscure any information from the source document.

6. De-NISTing and Deduplication. Electronic file collections will be De-NISTed, removing commercially available operating system and application file contained on the current NIST file list. Defendants will globally deduplicate identical ESI as follows:

a. Electronic Files: Duplicated electronic files will be identified based upon calculated MD5 Hash values for binary file content. File contents only will be used for MD5 Hash value calculation and will not include operating system metadata (filename, file dates) values. All files bearing an identical MD5 hash value are a duplicate group. The document reviewed by Defendants for privilege, relevance, or confidentiality shall be deemed the primary duplicate document within the group. Generally, the Defendants shall not remove any of the objective coding fields listed in paragraph G.1 above, in either primary or duplicate documents. If redactions are applied to the subject and/or text fields, however, Defendants may apply the

same redactions to all other documents within the duplicate group. Defendants shall only produce one document image or native file for duplicate ESI documents within the group. For Takeda sources, the following metadata fields as described in Section G.1 associated with the produced document will provide information for duplicate documents not produced: CustodianAppendMulti and RelativePathAppend.

b. Messaging Files: Duplicate messaging files will be identified based upon MD5 Hash values for the message family, including parent object and attachments. The following fields will be used to create the unique value for each message: To; From; CC; BCC; Date Sent; Subject; Body; and, MD5 Hash values for all attachments, in attachment order. Duplicate messaging materials will be identified at a family level, including message and attachment(s). All files bearing an identical MD5 Hash value are a duplicate group. The documents reviewed by Defendants for privilege, relevance, or confidentiality shall be deemed the primary duplicate document within the group. For identified duplicate ESI, the Defendants shall not remove any of the objective coding fields listed in paragraph G.1 above. If redactions have been applied to such fields, Defendants may substitute and replace the subject and text fields with those reviewed by Defendants' counsel for the primary duplicate ESI document for the other documents within the duplicate group. Defendants shall only produce one document image or native file for duplicate ESI documents within the group. For Takeda sources, the following metadata field as described in Section G.1 associated with the produced document will provide information for duplicate documents not produced: CustodianAppendMulti.

c. E-mail Threading: The producing Party may identify e-mail threads where all previous emails which make up the thread are present in the body of the final e-mail in the thread. Any party electing to use this procedure must notify all receiving parties that e-mail

thread suppression has been proposed to be performed on a specified production and the Parties agree to meet and confer regarding the format of this production, and reserve the right to seek Court guidance on the issue should agreement not be reached.

7. Production Media. The producing Party may produce documents via a secure file transfer mechanism and/or on readily accessible, computer or electronic media as the Parties may hereafter agree upon, including CD-ROM, DVD, external hard drive (with standard PC compatible interface), (the "Production Media"). Each piece of Production Media will be assigned a production number or other unique identifying label corresponding to the date of the production of documents on the Production Media (e.g., "Defendant Takeda Production April 1, 2012") as well as the sequence of the material in that production (e.g. "-001", "-002"). For example, if the production comprises document images on three DVDs, the producing Party may label each DVD in the following manner "Defendant Takeda Production April 1, 2012", "Defendant MSL Production April 1, 2012-002", "Defendant Takeda Production April 1, 2012-003." Additional information that will be identified on the physical Production Media includes: (1) text referencing that it was produced in *In re: Actos (Pioglitazone) Products Liability Litigation*; and (2) the Bates Number range of the materials contained on the Production Media. Further, any replacement Production Media will cross-reference the original Production Media and clearly identify that it is a replacement and cross-reference the Bates Number range that is being replaced.

8. Write Protection and Preservation. All computer media that is capable of write protection should be write-protected before production.

9. Inadvertent Disclosures. The terms of the Case Management Order: Assertions of Attorney-Client Privilege and Work Product Doctrine shall apply to this protocol.

10. Duplicate Production Not Required. The Parties shall meet and confer regarding any Party's request to produce identical paper copies of data already produced in electronic form.

H. Timing.

1. The Parties will use their reasonable efforts to produce ESI in a timely manner consistent with the Court's discovery schedule.

2. The Parties will produce ESI on a rolling basis.

I. General Provisions.

1. Any practice or procedure set forth herein may be varied by agreement of the Parties, and first will be confirmed in writing, where such variance is deemed appropriate to facilitate the timely and economical exchange of electronic data.

2. Should any Party subsequently determine it cannot in good faith proceed as required by this protocol; the Parties will meet and confer to resolve any dispute before seeking Court intervention.

3. The Parties agree that e-discovery will be conducted in phases and the Parties will meet and confer regarding discovery of data sources not listed herein.

4. Regardless of the foregoing, the Parties are under a continuing obligation to produce identified responsive, non-privileged documents and to identify sources of potentially discoverable materials consistent with their obligations under Federal Rules of Civil Procedure.

J. Items Requiring Meet and Confer.

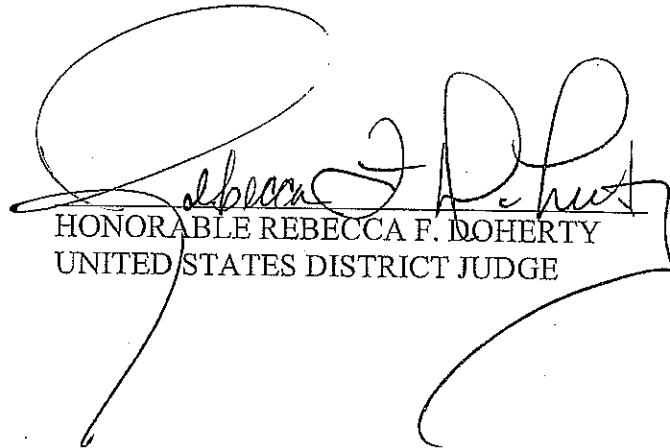
1. The Parties agree to meet and confer regarding the following items in advance of impacted productions:

- a. Whether the E-mail Property metadata field is able to be produced
- b. Certain technical specifications for productions:

- (1) Hard copy document unitization
- (2) Microsoft "Auto" features or macros
- (3) Embedded objects
- (4) Compressed Files
- (5) Load file organization

IT IS SO ORDERED.

THUS DONE AND SIGNED in Lafayette, Louisiana, this 27 day of July, 2012.



HONORABLE REBECCA F. DOHERTY
UNITED STATES DISTRICT JUDGE

EXHIBIT A

**UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF LOUISIANA**

In Re: Actos (Pioglitazone) Products)	MDL NO. 6:11-md-2299
Liability Litigation)	
)	JUDGE DOHERTY
)	
This Document Applies to:)	MAGISTRATE JUDGE HANNA
)	
All Cases)	
)	
)	
)	
)	

NONDISCLOSURE AND CONFIDENTIALITY AGREEMENT

I, _____, state the following:

1. I have read and I understand the Protocol Relating to the Production of Electronically Stored Information and Order in the above-captioned action (a copy of which is attached and whose definitions are incorporated herein) and I attest to my understanding that access to information of the Parties may be provided to me and that such access shall be governed by the terms and conditions and restrictions of the Order. I agree to be bound by the terms of the Order and hereby subject myself to the jurisdiction of the Court for all purposes related to the Order.

2. I shall not use or disclose any information to others, except in accordance with Order. I also understand that, in the event that I fail to abide by the terms of this Nondisclosure and Confidentiality Agreement and/or Order, I may be subject to sanctions by way of contempt of court and to separate legal and equitable recourse by the Parties.

3. I will comply with all of the provisions of the Order. I will hold in confidence, will not disclose to anyone not qualified under the Order, and will use only for the purposes authorized under the Order for the above-captioned litigation any information that is disclosed to me.

4. I understand that my agreement survives the termination of the above-captioned litigation, whether by settlement or other termination. I therefore understand such termination shall not relieve me from any of the continuing obligations of confidentiality imposed by the Order.

5. I hereby submit to the jurisdiction of the United States District Court to the extent allowed by law and to the full extent determined by the United States District Court for the purpose of enforcement of this Nondisclosure and Confidentiality Agreement and the Order.

Signature

Printed Name and Address

Date

EXHIBIT D

UNITED STATES DISTRICT COURT
DISTRICT OF MINNESOTA

IN RE: BAIR HUGGER FORCED AIR
WARMING PRODUCTS LIABILITY
LITIGATION

MDL No. 15-2666
(JNE/FLN)

This Document Relates to All Actions

**PRETRIAL ORDER NO. 12:
Computer Assisted Review Protocol**

PROTOCOL REGARDING COMPUTER ASSISTED REVIEW

Pursuant to the Parties' agreement, the procedures outlined herein shall govern the use of computer assisted review ("CAR") for electronically stored information ("ESI") during the pendency of this litigation.

I. Data Sources. To the extent either party collects e-mail or other electronic documents, they will do so for at least the data sources agreed upon with the opposing party. Non-custodial data sources, such as relevant databases or group shares, shall also be discussed. The Parties agree that the stipulations set forth in this Protocol shall not limit their ability to request discrete documents or document categories that may not belong to a particular agreed-upon custodian or data source. Nothing in this Protocol shall be construed to waive the Parties' rights to request documents or to object to particular discovery requests as they are entitled under the Federal Rules of Civil Procedure. The Parties will confer in good faith upon any such requests or objections. If the Parties cannot reach agreement, any dispute shall be presented to the Court in accordance with the Federal Rules of Civil Procedures, Local Rules, and orders of the Court.

II. Computer Assisted Review. The parties reserve the right to use CAR, also referred to "technology assisted review," or "predictive coding," in which a machine-

learning algorithm assesses the likely relevance of a corpus of documents based on manual relevance determinations applied to a subset of that corpus. Although the Parties presume that each party is in the best position to know how to search for its own responsive ESI that will be considered for CAR, the Parties have the right to make reasonable inquiries concerning the adequacy and success of the other's process. As with key word search term negotiation, the use of CAR is best when both parties can agree as to what documents constitute responsive versus non-responsive material. The goal here is to get to the facts and not be bogged down with technological minutia. The following protocol is designed so as not to be unduly burdensome, while also allowing both Parties to reap the benefits of using CAR, including greater accuracy than keyword search or manual review, reduced costs, and the more rapid document delivery time frames offered by CAR. To the extent the Protocol does not achieve these goals in practice, the Parties reserve their rights to meet-and-confer and seek appropriate modification. The protocol for the review and production of documents using CAR is as follows:

a. **Documentation of the CAR Startup Process.** Each party leveraging CAR will produce a written description, with reasonable levels of detail, of the method or methods used to search for documents and data responsive to the properly served requests for production, that will be considered for CAR, including but not limited to the custodial and non-custodial sources being searched, the rationale for that selection, the specific software being used for CAR, and the specifications being implemented.

Once this has been negotiated and agreed to as being satisfactory, CAR seed set training will commence.

b. **Training Methodology.** At producing party's sole option, but subject to receiving party's right to raise this issue with the Court per paragraph II.g.4.iv if the producing party elects not to follow this procedure, the following procedure will be used to designate individuals who will be responsible for document assessment and generation of the CAR document training set. Each party will designate two subject-matter experts. These designated persons will work together at the same time, in the same room, to determine which documents are either responsive or non-responsive. These designated persons are not allowed any form of phone, computer, tablet, and paper or information recording device. What is seen in the room, stays in the room. Furthermore, these experts shall agree in writing not to disclose the content of nonresponsive or privileged documents to anyone other than the producing party and will be subject to sanction for any disclosure or improper use. Whether or not the producing party elects this procedure, the remainder of this protocol shall be used by the producing party in the performance of CAR.

- (1) **Initial CAR Training:** Building the initial training model for the computer can be a time consuming process, while subsequent training across newly added documents takes less effort. For the producing party's initial model build: the Parties may each identify up to 250 unique production documents to be included in the initial

feed into the system to speed up training. In addition to the production documents above, no less than 1500 randomly selected documents will be reviewed and categorized by the producing party for initial training.

c. **Privileged Documents.** Documents that are privileged are often good training materials for the CAR system. If a document, at first glance, is determined to be privileged, the experts from the opposing party, if they are in the room, are to look away, while the privilege content is assessed solely by the producing party. If this content contains fact patterns outside of this matter, the document should be excluded, otherwise if relevant, it should be included. Post-CAR-training, these privileged documents can be filtered and removed from the final production. Training the system as outlined above using privileged documents does not result in any disclosure of the privileged document to the opposing party and does not constitute a waiver of privilege.

d. **Documents Not Suitable for CAR.** Documents that are not appropriate for CAR retrieval, including but not limited to spreadsheets, databases, flat-file PDF documents, Computer-Aided-Design files, video, audio, and images, will be not be subject to the CAR process, but may still be requested to be produced in response to properly served requests for production.

e. **Training Targets.** Once adequate training of the system has been achieved, the CAR process will categorize the target documents. Adequate training of the CAR system shall be reached when both Recall rate is at or exceeds 80% and the

producing party determines in good faith that the burden of additional enhancement of Recall Rate outweighs the benefit. If the producing party is unable to meet the 80% threshold for Recall with reasonable effort, the Parties shall meet and confer on other steps that can reasonably be taken to meet that threshold and/or modification of the threshold if necessary. If the Parties cannot agree on other steps and/or modification of the threshold, the Parties shall bring the issue to the Court for resolution.

f. **Use of Exemplars and Remediation.** The Parties can also elect to use certain exemplars of good or bad documents obtained through keyword searching or other means to adjust the training of the CAR system so that its overall accuracy is improved at any time during the process. However, keyword search, concept search, or other similar methodologies will not be used prior to predictive coding being run, unless and until, for a specific collection of documents, the rate of prevalence is too low to achieve the targeted level of performance via normal means. This procedure contains a number of remediation steps that can be taken prior to the need to perform such actions and these are to be followed until such time as to be deemed ineffective. If the producing party believes keyword search, concept search or other similar methodologies has become necessary, the Parties will meet and confer regarding the use of such methodology, and if no agreement can be reached, the issue shall be brought to the Court for resolution.

g. **Measurements and Documentation.** The following pre- and post-run information will be captured, documented, and shared for each and every run in order

to determine that the CAR training was appropriately carried out and in order to determine that Recall rate in II.e is being met:

- (1) A written description of the methods that were used to assess the adequacy of the process used.
- (2) The following assessment information:
 - i. The number of documents (or other ESI units) that were collected by the producing party and the means used for identifying those documents.
 - ii. The steps that were taken to evaluate the success of any method, other than the CAR contemplated herein, used to reduce the number of documents collected to hold down the number that was ultimately submitted to the CAR system (culling evaluation).
 - iii. The number and percentage of total documents submitted to the CAR system as the post-culling population.
 - iv. The number and percentage of the documents that were ultimately deemed responsive, exclusive of their non-responsive family members. That is, the number of documents that were actually identified by the CAR system as responsive.
 - v. The levels of Recall and Precision that were achieved, and how these were calculated. These will be reported both pre-run and post-run, but prior to audit.

(3) **Audit.** The results of each predictive coding run will be audited using the following process:

- i. A random sample of 1,000 documents drawn from the set of documents, post-categorization, that were not considered potentially responsive and therefore not produced (called the “negative set”).
- ii. A similar random sample will also be produced from the responsive set.
- iii. The method used to select the random documents will be described by producing party detailing the technical methodology and/or workflow used to create the sample document sets.
- iv. Both sets of randomly selected documents will be audited for accuracy, relative to the set of documents they represent. The persons who will perform the audit shall be the same individuals who conducted the initial training.
- v. Prior to this audit occurring, the documents will be rapidly assessed for privilege solely by the producing party and those privileged documents withheld from confirmatory review. These privileged documents will still be included in the overall accuracy count for purposes of assessment. The objective here is

to determine the effectiveness of the CAR training model and its overall performance.

- vi. The goal of the audit is to assess if potentially responsive materials were located in the negative set, and if responsive materials contain content that should be in the negative set. The goal of the measurement is to determine the overall frequency of such wrong designation, and to assess whether this is acceptable or otherwise based upon agreed to performance metrics and thresholds above.
- vii. Up to 400 document moves can occur per audit (irrelevant to relevant and vice-versa). Once these changes have occurred, predictive coding can be re-run and accuracy re-assessed.
- viii. The number of documents moved and in which direction they moved will be documented.
- ix. The performance ratios obtained through this measurement process will be documented, capturing the effective precision and recall rates, as well as the number of privileged documents that were withheld from confirmatory review.
- x. If the recall rate agreed to above is not obtained, the process will iterate until it is.

- xi. For the initial set of documents used for training, and for each pass through this QC process promptly after each pass being completed, regardless of whether the targeted recall rate is obtained, not obtained, or exceeded, the producing party will make available to the receiving party a log in Excel format containing:
 - a. the CAR run date & time;
 - b. a list of control numbers for all documents the system identified as responsive, along with a copy of those documents except that no copy need be provided for those documents withheld for privilege;
 - c. a list of control numbers for all documents the system identified as non-responsive, along with a copy of those documents the system incorrectly identified as non-responsive and which were moved from the non-responsive set to the responsive set, except that no copy need be provided for those documents withheld for privilege;
 - d. a count of the number of privileged documents correctly identified as responsive but withheld for privilege;
 - e. a count of the number of privileged documents incorrectly identified as responsive but withheld for privilege;

- f. the metadata for all documents as described in Pretrial Order No. 10 at paragraph II.P or Pretrial No. 11 depending on the privilege status of the document;
- g. four additional columns in the spreadsheet indicating: (1) whether the document was marked by the CAR system as responsive or non-responsive; (2) whether, after manual review, the document was changed from the non-responsive set to the responsive or vice-versa; (3) if a copy of the document is not produced by reason of privilege, the reason for withholding the document on the basis of privilege; and 4) for any document for which a copy is not being produced due to non-responsiveness or privilege, information sufficient to allow the receiving party to determine why the document was considered non-responsive and/or privileged.¹
- xii. The receiving party has fourteen (14) calendar days to review the log and associated production and meet and confer to request or suggest changes and, if no agreement can be reached, bring the issue to the Court for resolution. The log shall be treated as

¹ Providing the level of detail required for privilege logs by Pretrial Order No. 11 satisfies this provision for privileged documents.

Confidential under the Protective Order and shall not be used for any purpose other than as provided by this Protocol.

- xiii. If during the audit it is found that a certain discrete class or sub-grouping of documents are being incorrectly selected by the predictive coding system, it is permissible to filter for these and remove these from the initial document population prior to predictive coding be re-run. If the producing party believes this is necessary, however, the Parties shall first meet & confer as to the reasons for the filter becoming necessary and the filtering process, and if no agreement can be reached, bring the issue to the Court for resolution.

(4) **Audit Assessment.** Based on the receiving party's evaluation of the information the providing party has produced under the paragraphs 2 & 3 above and for each production of documents, the receiving party may, at its election, do one or more of the following:

- i. Accept the use of CAR as is;
- ii. Request a meet-and-confer to discuss additional custodians, time ranges, key words or other means for identifying new documents to be considered for responsiveness;
- iii. Request a meet-and-confer to discuss modifications to the CAR process; and/or

- iv. Ask for Court intervention if the Parties cannot agree to an appropriate resolution through steps i-iii above, including but not limited to asking the Court to allow the receiving party to participate in the training under paragraph II.b and audit of the CAR system under paragraph II.g.3.
- (5) The Parties will attempt in good faith to resolve any disagreements through a meet and confer process. Should the Parties be unable to reach agreement, the dispute will be submitted to the Court for resolution.
- (6) CAR is an iterative process. As new document sources are selected for review or additional information is obtained by the Parties, additional training and ongoing assessment of the CAR system will be performed using the protocols set out above.

IT IS SO ORDERED.

Date: July 8, 2016

s/ Joan N. Ericksen
JOAN N. ERICKSEN
United States District Judge

EXHIBIT E

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF INDIANA
SOUTH BEND DIVISION

IN RE: BIOMET M2a MAGNUM HIP
IMPLANT PRODUCTS LIABILITY
LITIGATION (MDL 2391)

CAUSE NO. 3:12-MD-2391

This Document Relates to All Cases

MEMORANDUM AND ORDER

Biomet had begun searching its electronic files before the creation of this multidistrict litigation docket through a keyword culling and de-duplication process, winnowing the universe of potentially relevant documents (the universe as Biomet saw it) from 19.5 million documents to 2.9 million. At that point, Biomet turned to technology-assisted review, or “predictive coding,” to winnow the potentially discoverable documents down further. On April 18, I resolved the parties’ dispute over whether Biomet had to go back to square one in its document production and use predictive coding. I ruled that Biomet didn’t have to do that, though I encouraged Biomet to do what it said it would do and let the Plaintiffs’ Steering Committee suggest additional search terms.

Today’s dispute picks up where the April order left off. The Steering Committee wants Biomet to produce the discoverable documents used in the training of the “predictive coding” algorithm. Biomet reveals only that the discoverable documents used in the seed set already have been disclosed to the Steering Committee; Biomet won’t identify the seed set beyond that. Without knowing the training documents, the Steering Committee says, it can’t intelligently propose more search terms, since it doesn’t know what already has

been included in the search. Biomet says nothing in the law requires it to provide what the Steering Committee seeks.

My understanding of the predictive coding process consists largely of what was placed before me in resolution of the first discovery issue, although Biomet added a little more information and terminology in the memoranda I requested from both sides. As I understand it, a predictive coding algorithm offers up a document, and the user tells the algorithm to find more like that document or that the user doesn't want more documents like what was offered up.

The Steering Committee wants the whole seed set Biomet used for the algorithm's initial training. That request reaches well beyond the scope of any permissible discovery by seeking irrelevant or privileged documents used to tell the algorithm what not to find. That the Steering Committee has no right to discover irrelevant or privileged documents seems self-evident.

So I might have misunderstood the Steering Committee's request. It might be that the Steering Committee only wants to know which discoverable documents were used in the seed set, placing the parties' disagreement in a different light. Biomet tells the court — and the Steering Committee is in no position to agree or disagree — that it already has produced all of the discoverable documents that were used in the seed set. Given that production, then, the Steering Committee doesn't seek the production of documents; it has all the documents it wants to know about. The Steering Committee wants Biomet to tell how it went about identifying and selecting the documents (not just the seed set) that it has produced.

I'm puzzled as to the authority behind the Steering Committee's request. Federal Rule of Civil Procedure 26(b)(1) makes discoverable

any nonprivileged matter that is relevant to any party's claim or defense—including the existence, description, nature, custody, condition, and location of any documents or other tangible things and the identity and location of persons who know of any discoverable matter. For good cause, the court may order discovery of any matter relevant to the subject matter involved in the action. Relevant information need not be admissible at the trial if the discovery appears reasonably calculated to lead to the discovery of admissible evidence. All discovery is subject to the limitations imposed by Rule 26(b)(2)(C).

The Steering Committee knows of the existence and location of each discoverable document Biomet used in the seed set: those documents have been disclosed to the Steering Committee. The Steering Committee wants to know, not whether a document exists or where it is, but rather how Biomet used certain documents before disclosing them. Rule 26(b)(1) doesn't make such information disclosable.

The only authority the Steering Committee cites is a report of the Sedona Conference that has had a significant, salutary, and persuasive impact on federal discovery practice in the age of electronically stored information. Sedona Conference Cooperation Proclamation, 10 Sedona Conf. J. 331 (Fall Supp. 2009). Biomet, the Steering Committee says, isn't proceeding in the cooperative spirit endorsed by the Sedona Conference and the corresponding Seventh Circuit project. But neither the Sedona Conference nor the Seventh Circuit project expands a federal district court's powers, so they can't provide me with authority to compel discovery of information not made discoverable by the Federal Rules.

Still, Biomet's position is troubling. Biomet suggests no way in which telling the Steering Committee which of the documents already produced were in the seed set would harm it. Based on what I have been given in the parties' memoranda, Biomet is right that it doesn't have to identify the seed set, but the Steering Committee is right that Biomet's cooperation falls below what the Sedona

Conference endorses. An unexplained lack of cooperation in discovery can lead a court to question why the uncooperative party is hiding something, and such questions can affect the exercise of discretion.

But I don't have any discretion in this dispute. I won't order Biomet to reveal which of the documents it has disclosed were used in the seed set, but I urge Biomet to re-think its refusal.

ENTERED: August 21, 2013

/s/ Robert L. Miller, Jr.
Robert L. Miller, Jr., Judge
United States District Court

EXHIBIT F

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

IN RE: VIAGRA (SILDENAFIL CITRATE)
PRODUCTS LIABILITY LITIGATION

Master File No.: 3:16-md-02691-RS

MDL No. 2691

This Document Relates to: ALL ACTIONS

**STIPULATED ORDER RE:
DISCOVERY OF ELECTRONICALLY
STORED INFORMATION**

1. PURPOSE

This Order will govern discovery of electronically stored information (“ESI”) in this case as a supplement to the Federal Rules of Civil Procedure, this Court’s Guidelines for the Discovery of Electronically Stored Information, and any other applicable orders and rules. This Order shall apply to the production of hard-copy and electronic documents by Pfizer Inc. and its agents, employees (current and former), representatives, subsidiaries, and other affiliated entities (collectively, “Pfizer”), as well as to the production of hard-copy and electronic documents by Plaintiffs. This Order also may apply to state court actions provided that the parties thereto so agree or the applicable court so orders.

Nothing in this Order alters a Party’s rights, obligations, and responsibilities under the Federal Rules of Civil Procedure and any other applicable orders and rules, nor does anything in this Order impose additional burdens beyond those imposed by the Federal Rules of Civil Procedure or any other applicable orders or rules. The Parties reserve all objections under the Federal Rules of Civil Procedure and applicable decision authority for matters relating to the production of documents that are not specifically addressed in this Order.

2. DEFINITIONS

(a) **“Confidentiality Designation”** means the legend affixed to Documents for Confidential or Highly Confidential Information as defined by, and subject to, the terms of Stipulated Protective Order entered by the Court in this litigation.

(b) **“Document”** is defined to be synonymous in meaning and equal in scope to the usage of this term in Rules 26 and 34 of the Federal Rules of Civil Procedure. The term “document” shall include hard-copy documents, electronic documents, and ESI as defined herein.

(c) **“Electronic Document or Data”** means documents or data existing in electronic form at the time of collection, including but not limited to: e-mail or other means of electronic communications, word processing files (e.g., Microsoft Word), computer presentations (e.g., PowerPoint slides), spreadsheets (e.g., Excel), and image files (e.g., PDF).

(d) **“Electronically stored information”** or **“ESI,”** as used herein has the same meaning as in Federal Rules of Civil Procedure 26 and 34.

(e) **“Hard-Copy Document”** means documents existing in paper form at the time of collection.

(f) **“Hash Value”** is a unique numerical identifier that can be assigned to a file, a group of files, or a portion of a file, based on a standard mathematical algorithm applied to the characteristics of the data set. The most commonly used algorithms, known as MD5 and SHA, will generate numerical values so distinctive that the chance that any two data sets will have the same Hash Value, no matter how similar they appear, is less than one in one billion.

(g) **“Load files”** means electronic files provided with a production set of documents and images used to load that production set into a receiving party’s document review platform, and correlate its data within that platform.

(h) **“Media”** means an object or device, real or virtual, including but not limited to a disc, tape, computer, or other device on which data is or was stored.

(i) **“Metadata”** means: (i) information embedded in or associated with a native file that is not ordinarily viewable or printable from the application that generated, edited, or modified such native file which describes the characteristics, origins, usage, and/or validity of the electronic file; (ii) information generated automatically by the operation of a computer or other information technology system when a native file is created, modified, transmitted, deleted, or otherwise manipulated by a user of such system, (iii) information, such as Bates numbers, created

1 during the course of processing documents or ESI for production, and (iv) information collected
 2 during the course of collecting documents or ESI, such as the name of the media device on which
 3 it was stored, or the custodian or non-custodial data source from which it was collected.

4 (j) **“Native Format”** means and refers to the format of ESI in which it was generated
 5 and/or as used by the Producing Party in the usual course of its business and in its regularly
 6 conducted activities. For example, the native format of an Excel workbook is a .xls or .xlsx file.

7 (k) **“Optical Character Recognition” or “OCR”** means the process of recognizing
 8 and creating a file containing, visible text within an image.

9 (l) **“Party” or “Parties”** means or refers to the named Plaintiffs and/or Pfizer in
 10 the above-captioned matter, as well as any later added plaintiffs or defendants.

11 (m) **“Searchable Text”** means the native text extracted from an electronic document
 12 and any Optical Character Recognition text (“OCR text”) generated from a hard-copy document
 13 or electronic image.

14 (n) **“Include” and “Including”** shall be construed to mean “include but not be
 15 limited to” and “including, but not limited to”.

16 **3. COOPERATION**

17 The Parties are aware of the importance the Court places on cooperation and commit to
 18 cooperate in good faith throughout the matter consistent with this Court’s Guidelines for the
 19 Discovery of ESI. The Parties shall meet and confer in good faith on any issue regarding ESI, as
 20 necessary, including any relating to custodians and data sources, that arise under this Order. In
 21 the event the Parties cannot reach an agreement on a disputed matter, the Parties shall submit the
 22 matter to the Court in accord with local rules.

23 **4. LIAISON**

24 The Parties have identified liaisons to each other who are and will be knowledgeable
 25 about and responsible for discussing their respective Party’s ESI efforts. Each e-discovery liaison
 26 will be, or have access to those who are, knowledgeable about the technical aspects of the Party’s
 27 e-discovery systems, including the location, nature, accessibility, format, collection, search
 28

1 methodologies, and production of ESI in this matter. The Parties will rely on the liaisons, as
 2 needed, to confer about ESI and to help resolve disputes without court intervention.

3 **5. PRESERVATION**

4 The Parties represent that they have issued litigation hold notices and taken reasonable
 5 steps to preserve data in this litigation. The Parties shall maintain, preserve, and not render less
 6 reasonably accessible documents which may contain responsive data, or are produced pursuant to
 7 this Order and/or in response to requests for production of documents.

8 **6. IDENTIFICATION OF DOCUMENTS AND ESI**

9 The Parties agree to meet and confer to discuss (a) the identification of custodial and non-
 10 custodial data sources, and (b) additional parameters for scoping the review and production
 11 efforts (e.g., application of date ranges, de-NIST'ing, etc.).

12 **[Plaintiffs' Position: The Parties shall use technology-assisted review ("TAR") to**
 13 **identify responsive documents and ESI. The Producing Party will select the TAR**
 14 **technology and methodology to be used on its documents and ESI and disclose the name of**
 15 **the technology to the Receiving Party. The parties will meet and confer to reach agreement**
 16 **on the manual classifications of documents used in the TAR process (including, as**
 17 **applicable, seed set, control set, and validation set documents); the parameters (including**
 18 **cut-off scores) used to deem the process complete; and all manual steps in the validation of**
 19 **that process. If the Parties cannot reach agreement, the matter will be submitted to the**
 20 **Court. No methodology, parameters, search terms, TAR, or other filtering technology or**
 21 **methodology shall be used by the Producing Party without notice to the Receiving Party. If**
 22 **requested by the Receiving Party, the Parties will meet and confer to discuss and resolve**
 23 **any issues related thereto. In advance of the meet and confer, the Parties will exchange**
 24 **sufficient information about the methodology, parameters, search terms, or TAR to enable**
 25 **possible resolution.]**

26 **[Defendant's Position: With respect to custodial files, the Parties will (a) identify**
 27 **and select custodians most likely to possess relevant documents pursuant to any Pretrial**
 28 **Orders entered by the Court governing discovery; and (b) apply agreed-upon search terms**

1 to those custodians' data sources. The Parties shall first exchange search terms, then meet
 2 and confer to discuss the search terms and the proper methodology for validating those
 3 search terms (such as sampling of documents that do not hit on the search terms). The
 4 selection of search terms may need to be iterative.]

5 The Producing Party will review all hard-copy and electronic documents that [**Plaintiffs'**
 6 **Position: are deemed to be responsive by the TAR process** / **Defendant's Position: contain**
 7 **any agreed-upon search terms**] for responsiveness and privilege prior to production. The fact
 8 that a document may have been [**Plaintiffs' Position: deemed to be responsive by the TAR**
 9 **process** / **Defendant's Position: retrieved by application of any agreed-upon search terms**]
 10 shall not prevent the Producing Party from withholding the document for lack of responsiveness
 11 or privilege.

12 No specific document of which the Producing Party is aware and knows to be responsive
 13 shall be withheld from production because it was not identified as responsive by the agreed
 14 methodologies or was not within an identifiable data repository or custodial or non-custodial data
 15 source.

16 7. PRODUCTION FORMAT AND PROCESSING SPECIFICATIONS

17 (a) **Production Format.** Unless otherwise specified in Section 7(b) or pursuant to
 18 Section 7(k) below, the Parties shall produce all documents in black-and-white, single page, 300
 19 DPI, tagged image file format ("TIFF") images, utilizing Group IV compression, with
 20 corresponding extracted full text and, to the extent possible, applicable metadata as specified in
 21 Exhibit A. Image file names will be identical to the corresponding Bates numbered images, with
 22 a ".tif" file extension. The Producing Party will brand all TIFF images in the lower right-hand
 23 corner with its corresponding Bates number, using a consistent font type and size, to the extent
 24 possible. The Bates number must not obscure any part of the underlying image. If the placement
 25 in the lower right-hand corner will result in obscuring the underlying image, the Bates number
 26 should be placed as near to the position as possible while preserving the underlying image.

27 (b) **Native Format.** The Parties shall produce Excel spreadsheets, audio files, and
 28 video files in native format, unless redacted, with applicable metadata as specified in Exhibit A

1 and extracted searchable text. If production in native format is necessary to decipher the
2 meaning, context, or content of a Word or PowerPoint document produced in TIFF, the Producing
3 Party will honor reasonable requests made in good faith for the production of specific documents
4 in native format.

5 (c) **Embedded Objects.** If documents contain embedded objects, the Parties shall
6 extract the embedded objects as separate documents and treat them like attachments to the
7 document. To the extent reasonably possible, images embedded in emails shall not be extracted
8 and produced separately.

9 (d) **Load Files.** Every document referenced in a production image load file must have
10 all corresponding images, text, and metadata. The name of the image load file must mirror the
11 name of the delivery volume and should have a .LFP, .OPT, or .DII extension. The volume
12 names must be reasonably consecutive (e.g., ABC001, ABC002). The load file must contain one
13 line per image. Every image in the delivery volume must be contained in the image load file.
14 The image key must be named the same as the Bates number of the image. Load files must not
15 span across media.

16 (e) **Foreign Language Documents.** Hard-copy documents and ESI that contain
17 languages other than English, in whole or in part, shall be produced in the original language(s),
18 along with all existing translations of the searchable text maintained in the ordinary course of
19 business.

20 (f) **Text Files.** A single text file shall be provided for each document. The text file
21 name shall be the same as the Bates number of the first page of the document to which it
22 corresponds. Files names shall not have any special characters or embedded spaces. Electronic
23 text must be extracted directly from the native electronic file unless the document requires
24 redaction, is an image file, or is any other native electronic file that does not contain text to
25 extract (e.g., non-searchable PDFs). In these instances, a text file will be created using OCR and
26 will be produced in lieu of extracted text. Except in the case of redacted documents, the
27 Receiving Party will not be required to rely upon a less accurate version of the text than the
28 Producing Party.

1 (g) **TIFFs of ESI.** TIFFs of ESI shall convey the same information and image as the
2 original document, including all non-redacted elements and formatting which are visible in any
3 view of the document in its native application.

4 (h) **Bates Numbers.** All Bates numbers will consist of a three digit Alpha Prefix,
5 followed immediately by an 8 digit numeric: AAA#####. There must be no spaces in the
6 Bates number. Any numbers with less than 8 digits will be front padded with zeros to reach the
7 required 8 digits.

8 (i) **Metadata Fields and Processing.** Each of the metadata fields set forth in Exhibit
9 A that can be reasonably extracted from ESI will be produced for each document. If a Party
10 becomes aware of a systemic issue extracting or processing metadata, the Party shall notify all
11 other Parties and they shall meet and confer to arrive at a mutually acceptable resolution of the
12 issue. The Parties are not obligated to populate manually any of the metadata fields in Exhibit A
13 if such fields cannot be extracted from a document.

14 (j) **Native File Image Placeholders.** A Bates-stamped placeholder TIFF, bearing the
15 legend "This document has been produced in native format" shall be provided for ESI produced
16 in native format; these placeholders will be Bates numbered in the same way as any other TIFF,
17 and the Bates number of that single page shall be used as the BegBates and EndBates of the
18 associated document.

19 (k) **Databases, Structured, Aggregated or Application Data.** The Parties will meet
20 and confer to address the production and production format of any responsive data contained in a
21 database or other structured or aggregated data source.

22 (l) **Scanning of Hard-Copy Documents.** The Parties may produce hard-copy
23 documents either in their hard-copy form or as scanned images. In scanning paper documents,
24 documents are to be produced as they are kept. For documents found in folders or other
25 containers with labels, tabs, or other identifying information, such labels and tabs shall be
26 scanned where practicable. The Parties will use best efforts to unitize documents.

27 (m) **Proprietary Software.** To the extent that relevant ESI cannot be rendered or
28 reviewed without the use of proprietary software, the parties shall meet and confer to ensure that

1 the data is produced in a format, or made accessible in a manner, that does not restrict the
2 receiving party's ability to utilize the data fully and to minimize any expense or burden associated
3 with the production of such documents in such format or access to such data in such manner.

4 (n) **Confidentiality Treatment.** The Parties have entered into a Stipulated Protective
5 Order in this matter, which specifies various confidentiality treatment levels for use in this matter.
6 The Producing Party will brand any confidentiality endorsements in a corner of any TIFF images
7 representing the produced item. Those endorsements must be in a consistent font type and size
8 and must not obscure any part of the underlying image or Bates number, to the extent possible.

9 (o) **Redactions.** A Party may use redactions to protect attorney-client or attorney
10 work product privileges, or consistent with the Protective Order entered in this matter. Other than
11 as allowed by the Stipulated Protective Order, no redactions for relevance may be made within a
12 produced document or ESI item. For redacted items which were originally ESI, unaffected, non-
13 privileged metadata fields will be provided and will include all non-redacted data. The basis for
14 each redaction must be provided as metadata, except for documents already produced as of the
15 date of this Order.

16 (p) **Color.** The Parties shall honor reasonable and specific requests for the production
17 of documents as color images.

18 (q) **Parent-Child Relationships.** Parent-child relationships (the association between
19 an attachment and its parent document or between embedded documents and their parent) shall be
20 preserved.

21 (r) **Family Relationships.** Family relationships often exist between an e-mail and its
22 attachments, but can also be found amongst a stand-alone document and files originally contained
23 within the parent document, which are subsequently de-embedded as part of discovery
24 processing. Non-relevant attachments may be excluded from production. All non-relevant
25 attachments excluded from production shall be produced as a slipsheet or placeholder.

26 (s) **OCR.** OCR software should be set to the highest quality setting during processing.
27 Documents containing foreign language text will be OCR'ed using the appropriate settings for
28

that language, e.g., OCR of German documents will use settings that properly capture umlauts. Settings such as “auto-skewing” and “auto-rotation” should be turned on during the OCR process.

(t) **Date Fields Time Zone.** All documents shall be processed so as to show fielded dates and times in UTC.

(u) **Explanation of Inability to Produce Metadata.** If the Producing Party is unable to produce metadata for a particular field or ESI document, the Parties shall then meet and confer to attempt to resolve the problem.

(v) **Lost, Destroyed or Irretrievable ESI.** If a Producing Party learns that responsive ESI that once existed was lost, destroyed, or is no longer retrievable as a result of acts or circumstances not occurring in the ordinary course of business, the Producing Party shall comply with its obligations under the Federal Rules of Civil Procedure.

8. PRODUCTION MEDIA

The Producing Party shall produce documents on readily accessible, computer or electronic media as the Parties may hereafter agree upon, including CD-ROM, DVD, external hard drive, or via secure FTP site (the “Production Media”). Each piece of Production Media shall be assigned a production number or other unique identifying label corresponding to the date of the production of documents on the Production Media as well as the sequence of the material in that production. The Producing Party shall accompany all document productions with a transmittal cover letter identifying by Bates number the documents produced.

9. PRIVILEGE LOG

Within sixty (60) days after production absent agreement of the Parties, the Producing Party shall provide the Receiving Party with a log of the documents entirely withheld from production for a claim of attorney-client privilege, work product protection, or other applicable privilege or immunity, as provided for under PTO __: Discovery and Other Proceedings Relating to General Causation. The log will be produced in Excel or CSV format, and populated with the following extracted metadata fields, to the extent providing this information will not destroy privilege: Custodian, From, To, CC, BCC, Subject, File Name, File Extension, File Path, Date Sent, Date Created, Date Last Modified. The log also will provide the

1 privilege(s) claimed. In-house attorney names shall be designated with an asterisk; outside
2 counsel attorney names will be designated with a double asterisk.

3 **10. INADVERTENT DISCLOSURE**

4 The Parties agree that they do not intend to disclose information subject to a claim of
5 attorney-client privilege or attorney work product protection. If, nevertheless, a Producing
6 Party inadvertently discloses privileged or work product information to another party, such
7 disclosure shall not constitute or be deemed a waiver or forfeiture of any claim of attorney-
8 client privilege or work product protection that the Producing Party would otherwise be entitled
9 to assert with respect to the inadvertently disclosed information and its subject matter.

10 If the Producing Party notifies the Receiving Party of the inadvertently disclosed
11 documents or information, the Receiving Party shall return or destroy, within fourteen (14)
12 business days, all copies of such documents or information and upon written request provide a
13 certification of counsel that all such inadvertently disclosed documents or information have
14 been returned or destroyed. After a Producing Party provides written notice of inadvertent
15 production, a Receiving Party shall not copy, distribute, or otherwise use in any manner the
16 inadvertently disclosed documents or information, and shall notify all persons to whom the
17 Receiving Party has disseminated a copy of the inadvertently disclosed documents or
18 information that the documents or information are subject to this Order and may not be copied,
19 distributed, or otherwise used pending further notice from the Court.

20 **12. COST SHIFTING**

21 Generally, the costs of production pursuant to this Order shall be borne by the Producing
22 Party. However, the Court may apportion the costs of electronic discovery in accordance with
23 the Federal Rules of Civil Procedure.
24
25
26
27
28

1 **13. MODIFICATION**

2 This Stipulated Order may be modified by a Stipulated Order of the Parties or by the
3 Court for good cause shown.

4 **IT IS SO STIPULATED**, through Counsel of Record.

5
6 Dated:

Counsel for Plaintiff

7
8 Dated:

Counsel for Defendant

9 **IT IS ORDERED** that the forgoing Agreement is approved.

10
11 Dated:

RICHARD SEEBORG
UNITED STATES DISTRICT JUDGE